



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification 5 :</b> C07H 21/00, C07K 13/00, 15/18 C12N 15/00, 15/11, 15/63 C12N 15/67, 15/70, 15/74 C12N 15/79, 15/85, C12P 21/02	<b>A1</b>	<b>(11) International Publication Number:</b> WO 93/13119  <b>(43) International Publication Date:</b> 8 July 1993 (08.07.93)
<b>(21) International Application Number:</b> PCT/US92/11194 <b>(22) International Filing Date:</b> 23 December 1992 (23.12.92)  <b>(30) Priority data:</b> 07/816,289                      3 January 1992 (03.01.92)      US 07/906,871                      2 July 1992 (02.07.92)              US  <b>(71) Applicant:</b> BRIGHAM AND WOMEN'S HOSPITAL [US/US]; 75 Francis Street, Boston, MA 02115 (US).  <b>(72) Inventors:</b> STEVENS, Richard, L. ; 99 Indian Ridge Road, Sudbury, MA 01776 (US). AVRAHAM, Shalom ; 50 Radnor, Apt. #2, Brighton, MA 02135 (US).		<b>(74) Agents:</b> FOX, Samuel, L. et al.; Sterne, Kessler, Goldstein & Fox, 1225 Connecticut Avenue, Suite 300, Washing- ton, DC 20036-2678 (US).  <b>(81) Designated States:</b> AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> HEMATOPOIETIC CELL SPECIFIC TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES THEREOF  <b>(57) Abstract</b>  A negative transcription regulatory element (a constitutive suppressor) between residues -250 and -190 of the 5' flanking region of the mouse serglycin gene, a positive (hematopoietic cell enhancer) regulatory element located between residues -118 and -81, an equivalent of the TATA- box and a novel eukaryotic promoter that utilizes such equivalent, vectors containing such elements and hosts transformed therewith, are described. These regulatory elements, vectors and hosts are useful in gene transcription of heterologous genes in eukaryotic cells, and especially in hematopoietic cells. In addition, transcriptional factors that bind to these elements are described.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CC	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LJ	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Monaco	TG	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

## TITLE OF THE INVENTION

HEMATOPOIETIC CELL SPECIFIC TRANSCRIPTIONAL  
REGULATORY ELEMENTS OF SERGLYCIN  
AND USES THEREOF5     Cross Reference to Related Applications

10     This application is a continuation-in-part of U.S. Application No. 07/816,289, filed January 3, 1992, which is a continuation-in-part of U.S. Application No. 07/635,544, filed January 18, 1991, which is the U.S. National Phase of PCT/US89/03051, filed July 13, 1989, which is a continuation-in-part of U.S. Application No. 07/224,035, filed July 13, 1988, now abandoned; the contents of each of these priority applications are fully incorporated herein by reference.

Statement of Government Interest

15     The research underlying this invention was supported with U.S. government funds; the U.S. government has certain rights in this invention.

Field of the Invention

20     The invention is in the area of recombinant DNA technology. Specifically, the invention is directed to a hematopoietic cell-specific transcriptional enhancer element, a transcriptional suppressor element, and a promoter element, all present in the 5' flanking region of the serglycin gene; the invention is further directed to recombinant vectors containing such elements, hosts transformed with such vectors, and the use of vectors and hosts for recombinant gene transcription. The invention  
25     is also directed to purified protein factors that specifically bind to the transcriptional elements of the invention.

### BACKGROUND OF THE INVENTION

The rat connective tissue mast cell was the first cell conclusively shown to store proteoglycans in an intracellular secretory granule compartment (Benditt *et al.*, *J. Histochem. Cytochem* 4:419 (1956)). Rat connective tissue mast cells contain up to 25 pg/cell of an acidically charged ~750 kDa (kilodalton) proteoglycan that possesses a very small peptide core to which approximately seven heparin glycosaminoglycans of 75-100 kDa are attached (Yurt *et al.*, *J. Biol. Chem.* 252:518 (1977); Robinson *et al.*, *J. Biol. Chem.* 253:6687 (1978); Metcalfe *et al.*, *J. Biol. Chem.* 255:11753 (1980)). Because the peptide core of mature rat heparin proteoglycan consists almost entirely of equal amounts of serine and glycine (Robinson *et al.*, *J. Biol. Chem.* 253:6687 (1978); Metcalfe *et al.*, *J. Biol. Chem.* 255:11753 (1980)) and because heparin glycosaminoglycan is O-glycosidically linked to serine at serine-glycine sequences within its peptide core (Lindahl *et al.*, *J. Biol. Chem.* 240:2817 (1965)), it was first postulated by Robinson and coworkers (Robinson *et al.*, *J. Biol. Chem.* 253:6687 (1978)) that the mature peptide core of this proteoglycan is predominantly an alternating sequence of serine and glycine.

It is now known that many cells of hematopoietic origin (including serosal mast cells, mucosal mast cells, basophils, natural killer cells, cytotoxic T lymphocytes, eosinophils, macrophages, and platelets) store a family of proteoglycans in a cytoplasmic granule compartment that is distinct from the plasma membrane-localized and extracellular matrix-localized families of proteoglycans (Stevens *et al.*, *Cur. Topics Microbiol. Immunol.* 140:93-108 (1988)). These intracellular proteoglycans (known as "serine-glycine rich proteoglycans," "SG-PG," "secretory granule proteoglycan," or "serglycin proteoglycans") have five to seven highly sulfated glycosaminoglycans attached O-glycosidically to a common 18,600 to 16,700  $M_r$  peptide core possessing a protease-resistant glycosaminoglycan attachment region that is a repeat of serine and glycine

amino acids (Yurt *et al.*, *J. Biol. Chem.* 252:518-521 (1977); Robinson *et al.*, *J. Biol. Chem.* 253:6687-6693 (1978); Razin *et al.*, *J. Biol. Chem.* 257:7229-7236 (1982); Stevens *et al.*, *J. Biol. Chem.* 260:14194-14200 (1985); Seldin *et al.*, *J. Biol. Chem.* 260:11131-11139 (1985); Bourdon *et al.*, *Proc. Natl. Acad. Sci. USA* 82:1321-1325 (1985); Bourdon *et al.*, *J. Biol. Chem.* 261:12534-12537 (1986); Avraham *et al.*, *J. Biol. Chem.* 263:7292-7296 (1988); Avraham *et al.*, *Proc. Natl. Acad. Sci.* 86:3763-3767 (1989); Stevens *et al.*, *J. Biol. Chem.* 263:7287-7291 (1988); Alliel *et al.*, *FEBS Lett.* 236:123-126 (1988); Stellrecht *et al.*, *Nuc. Acids Res.* 17:7523 (1989)). The peptide core of this family of proteoglycans has also been referred to by a variety of names, such as "secretory granule proteoglycan peptide core protein," but most recently has simply been called "serglycin." Thus, the gene encoding this peptide is the serglycin gene.

Serglycin proteoglycans (serglycin with attached glycosaminoglycans) are stored inside cells as a macromolecular complex bound to basically charged proteins. Because these proteoglycans are bound by ionic linkage in the secretory granules of mouse and rat mast cells to positively charged endopeptidases and exopeptidases that are enzymatically active at neutral pH, it has been assumed that the serglycin proteoglycans prevent intragranular autolysis of the proteases. The proteoglycan/protease macromolecular complexes remain intact when they are exocytosed from activated mast cells (Schwartz *et al.*, *J. Immunol.* 126:2071-2078 (1981); Serafin *et al.*, *J. Biol. Chem.* 261:15017-15021 (1986); Serafin *et al.*, *J. Immunol.* 139:3771-3776 (1987); Le Trong *et al.*, *Proc. Natl. Acad. Sci. USA* 84:364-367 (1987)), presumably attenuating diffusion of the proteases from inflammatory sites and facilitating concerted proteolysis of protein substrates.

cDNAs that encode serglycin have been isolated from rat (Bourdon *et al.*, *Proc. Natl. Acad. Sci. USA* 82:1321-1325 (1985); Bourdon *et al.*, *J. Biol. Chem.* 261:12534-12537 (1986); Avraham *et al.*, *J. Biol. Chem.* 263:7292-7296 (1988)), mouse (Avraham *et al.*, *Proc. Natl. Acad. Sci.*

86:3763-3767 (1989)), and human (Stevens *et al.*, *J. Biol. Chem.* 263:7287-7291 (1988); Alliel *et al.*, *FEBS Lett.* 236:123-126 (1988); Stellrecht *et al.*, *Nuc. Acids Res.* 17:7523 (1989)) cDNA libraries. These cDNAs encode 1.0-, 1.0-, and 1.3-kb transcripts in the mouse, rat, and human, respectively. The mouse serglycin gene resides on chromosome 10, is approximately 15 kb in size, and consists of three exons (Avraham *et al.*, *Proc. Natl. Acad. Sci.* 86:3763-3767 (1989)).

Bourdon and coworkers (Bourdon *et al.*, *Proc. Natl. Acad. Sci. USA* 82:1321 (1985); Bourdon *et al.*, *J. Biol. Chem.* 261:12534 (1986)) isolated and characterized a cDNA from a rat yolk sac tumor cell that encoded an unusual 18.6 kDa proteoglycan peptide core with a 49 amino acid glycosaminoglycan attachment region of alternating serine and glycine. Because of the preponderance of these two amino acids, it was proposed that the peptide core of this proteoglycan (designated serglycin) was related to the peptide core of rat mast cell-derived heparin proteoglycan. Numerous molecular biology studies have been carried out on the cDNAs and genes that encode mouse, rat, and human serglycin. Using a 3' gene-specific fragment of a rat serglycin cDNA (Avraham *et al.*, *J. Biol. Chem.* 263:7292 (1988)), it was demonstrated that this gene is expressed at relatively high levels in a variety of mouse and rat mast cells irrespective of what type of glycosaminoglycan is polymerized onto the peptide core (Tantravahi *et al.*, *Proc. Natl. Acad. Sci. USA* 83:9207 (1986)). This gene is also expressed in many other hematopoietic cells that possess secretory granules (Tantravahi *et al.*, *Proc. Natl. Acad. Sci. USA* 83:9207 (1986); Stevens *et al.*, *J. Immunol.* 139:863 (1987); Stevens *et al.*, *J. Biol. Chem.* 263:7287 (1988); Rothenberg, M. E., Pomerantz, J.L., Owen, W.F., Avraham, S., Soberman *et al.*, *J. Biol. Chem.* 263:13901 (1988); Stellrecht *et al.*, *Nucleic Acids Res.* 17:7523 (1989); Perin *et al.*, *Biochem. J.* 255:10017-1013 (1988); MacDermott *et al.*, *J. Exp. Med.* 162:1771 (1985); Nicodemus *et al.*, *J. Biol. Chem.* 265:5889 (1990)) and it appears that the same peptide core is used in all of these cell types. The selection of the

-5-

type of glycosaminoglycan that will be synthesized onto this peptide core therefore appears to be a cell-specific event that is not exclusively dependent on the translated peptide core.

5 Although serglycin is specifically expressed in hematopoietic cells, no tissue specific hematopoietic cell transcriptional regulatory elements have yet been identified. A need exists for such elements as they would allow, for the first time, the regulated induction or expression of recombinant genes in hematopoietic cells, especially in hematopoietic cell culture systems.

10

### SUMMARY OF THE INVENTION

Recognizing the importance of understanding tissue specific gene expression in hematopoietic cells for the expression of recombinant genes in cells of hematopoietic cell lineage, and cognizant of the need for DNA regulatory elements or motifs capable of specifically stimulating or inhibiting transcription for the controlled expression of genes in such cells, 15 the inventors investigated the 5' flanking region of the serglycin gene in an attempt to identify such motifs. These studies have culminated in the identification of three motifs in the 5' flanking region of the mouse serglycin gene that regulate the constitutive transcription of that gene.

20

According to the invention, there is first provided, in isolated form, a genetic sequence of approximately the proximal 500 nucleotides of the 5' flanking region of human and mouse serglycin gene, such flanking region providing transcriptional regulatory elements sufficient to direct expression of operably linked recombinant genes in hematopoietic host cells in a constitutive manner. 25

The invention further provides, in isolated form, genetic sequences encoding a positive transcriptional regulatory element, herein termed an enhancer element, such element corresponding to nucleotides -118 through -81 of the mouse serglycin gene (5 -

TCTGGGTGTTGATGTGGATCTCTTTCTATTGTTTCAGG-3' [SEQ ID No. 16]), and an equivalent position of the human serglycin gene, and such element being dominantly active to stimulate transcription of operably linked genes in hematopoietic host cells.

5       The invention further provides, in isolated form, genetic sequences encoding a unique and atypical eukaryotic promoter element, such promoter element corresponding to nucleotides -40 through -20 of the mouse serglycin gene (5'-GAACCTCTTTCTAAAAGGGAC-3' [SEQ ID No. 17], and an equivalent position of the human serglycin gene, and such  
10       element being dominantly active for the promotion of transcription in operably linked genes in hematopoietic host cells.

      The invention further provides, in isolated form, genetic sequences encoding a negative transcriptional regulatory element, herein termed a suppressor element, such element corresponding to nucleotides -250  
15       through -190 of the mouse serglycin gene (5'-TGCAAATGACAGATGGCAGA GCTTTTGGAAAAAGAAAAAA TAATAACCACACAGCAAACG-3' [SEQ ID No. 18], and an equivalent position of the human serglycin gene, and such element being dominantly active to inhibit transcription of operably linked genes in fibroblast host  
20       cells.

      The invention further provides expression vectors containing such genetic sequences, such expression vectors providing such genetic sequences in a manner that permits a gene of interest to be operably linked to the regulatory element encoded by the genetic sequence such  
25       that transcriptional expression of the gene of interest is under the control of the genetic sequence of the invention.

      The invention further provides expression vectors containing such genetic sequences operably linked to a gene of interest.

      The invention further provides host cells transformed with such  
30       expression vectors.



-7-

The invention further provides methods for the production of a peptide of interest, or for inhibiting the production of a peptide of interest, using the genetically engineered genetic sequences, vectors and hosts of the invention.

5       The invention further provides methods for the inhibition of the expression of a gene of interest, using the genetically engineered genetic sequences of the invention to direct the transcription of an anti-sense RNA complementary to the gene of interest.

10       The invention further provides cell-free preparations of B/F<sub>(-250/-161)</sub>-I, a *trans*-acting factor extractable from the nuclei of rat basophilic leukemia-1 cells and rat-1 fibroblasts, such factor specifically binding to the suppressor element of the invention.

15       The invention further provides cell-free preparations of B/F<sub>(-250/-161)</sub>-II, a *trans*-acting factor extractable from the nuclei of rat-1 fibroblasts, such factor specifically binding to the suppressor element of the invention.

The invention further provides cell-free preparations of B<sub>(-118/-81)</sub>-I, a *trans*-acting factor extractable from the nuclei of rat basophilic leukemia-1 cells, such factor specifically binding to the enhancer element of the invention.

20       The invention further provides cell-free preparations of F<sub>(-118/-81)</sub>-I, a *trans*-acting factor extractable from the nuclei of rat-1 fibroblast cells, such factor specifically binding to the enhancer element of the invention.

25       The invention further provides cell-free preparations of B/F<sub>(-40/+24)</sub>-I a *trans*-acting factor extractable from the nuclei of rat basophilic leukemia-1 cells and rat-1 fibroblast cells, and to F<sub>(-40/+24)</sub>-II, B<sub>(-40/+24)</sub>-II, *trans*-acting factors extractable from the nuclei of rat-1 fibroblast cells or rat basophilic leukemia-1 cells, respectively, such factors specifically binding to the enhancer element of the invention.

### BRIEF DESCRIPTION OF THE FIGURES

**Figure 1.** Restriction map and nucleotide sequencing strategy of cDNA-H4 and its related four cDNAs. The cDNA-H4 originated from HL-60 cells (a promyelocytic leukemia cell line). "X" and "A" refer to the sites within each cDNA which are susceptible to *Xmn*I and *Acc*I, respectively. The arrows indicate the direction and length of each subcloned fragment of cDNA that was sequenced.

**Figure 2.** Consensus nucleotide sequence of the HL-60 cell-derived cDNAs and the predicted amino acid sequence of the translated proteoglycan peptide core (serglycin) [SEQ ID Nos. 9 and 10]. The arrow indicates the putative site of cleavage of the signal peptide. Stop codons are indicated by \*\*\*. The number on the right and left indicate the amino acid and the nucleotide in the respective sequence. The *Xmn*I and *Acc*I restriction sites are indicated. The 5' end of the cDNA-H12 was 4 bp longer, and the 5' end of cDNA-H19 was 14 bp shorter than cDNA-H4. cDNA-H8 differed from the cDNA-H4 in that it had an extra thymidine (shown in parentheses) at the 3' end of its cDNA.

**Figure 3.** Restriction map of the human serglycin gene isolated from a human leukocyte genomic DNA library (Klickstein, L.B. *et al.*, *J. Exp. Med.* 165:1095-1112 (1987)).

**Figure 4.** A. Nucleotide sequence of the human serglycin gene [SEQ ID Nos. 11 (nucleotide sequence and 12 (protein sequence))]. The nucleotide sequences of the 5' flanking region, the exon/intron junctions, and the three exons are depicted. The hydrophobic signal peptide of the translated proteoglycan peptide core in exon 1 and the serine-glycine rich glycosaminoglycan attachment region in exon 3 are boxed. The polyadenylation site in exon 3 is underlined.

B. The complete nucleotide sequence of the human serglycin gene, including introns [SEQ ID No. 15].

**Figure 5.** Nucleotide sequence of the mouse serglycin gene [SEQ ID Nos. 13 (nucleotide sequence and 14 (protein sequence)], Avraham, S. *et al.*, *J. Biol. Chem.* 264:16719-16726 (1989). The nucleotide sequence of the 5' flanking region, the exon/intron junctions, and the three exons are depicted. The arrow indicates the probable transcription-initiation site. The hydrophobic signal peptide of the translated proteoglycan peptide core is boxed in exon 1. The di-acidic amino acid sequence that has been proposed to dictate glycosaminoglycan addition to proteins and the serine-glycine rich, glycosaminoglycan attachment region are boxed in exon 3. The polyadenylation site in exon 3 is underlined.

**Figure 6.** ELISA of the rabbit anti-peptide 02 serum. Peptides 01 (o--o) and 02 (●--●) were coupled to separate microtiter wells and different dilutions of the rabbit anti-peptide 02 sera were examined for their reactivity against the specific peptide as detected spectrophotometrically at 490 nm after the addition of horseradish peroxidase-conjugated goat anti-rabbit antibody followed by 2,2'-azino-di-[3-ethyl-benzthiazoline] sulfonate. The amino acid sequence of peptide 01 is Ser-Val-Gln-Gly-Tyr-Pro-Thr-Gln-Arg-Ala-Arg-Tyr-Gln-Trp-Val-Arg. The amino acid sequence of peptide 02 is Ser-Asn-Lys-Ile-Pro-Arg-Leu-Arg-Thr-Asp-Leu-Phe-Pro-Lys-Thr-Arg.

**Figure 7.** SDS-PAGE analysis of immunoprecipitates of lysates of [<sup>35</sup>S]sulfate-labeled (A) and [<sup>35</sup>S]methionine-labeled (B) HL-60 cells. (A) Lysates of [<sup>35</sup>S]sulfate-labeled HL-60 cells were analyzed before (lane 1) and after immunoprecipitation with anti-peptide 02 IgG in the presence (lane 2) or absence (lane 3) of peptide 02. (B) Lysates of HL-60 cells were analyzed after a 2 min (lane 1) or a 10 min (lane 2) incubation with [<sup>35</sup>S]methionine by immunoprecipitation with anti-peptide 02 IgG. Ten min [<sup>35</sup>S]methionine-labeled HL-60 cells were washed and then incubated for an additional 5 min in methionine-containing enriched medium before lysates were immunoprecipitated with anti-peptide 02 IgG (lane 3). Lysates of 5 min labeled HL-60 cells were immunoprecipitated with anti-

-10-

peptide 02 IgG in the presence of peptide 01 (lane 4) or peptide 02 (lane 5). The [<sup>35</sup>S]methionine-labeled proteins that were nonspecifically immunoprecipitated with pre-immune IgG are depicted in lane 6. The origin (ori) and the M<sub>r</sub> markers are indicated on the far left and right of each panel. The arrows indicate the precursor peptide core and the mature proteoglycan that are immunoprecipitated with anti-peptide 02 IgG.

**Figure 8. A.** Effects of progressive deletion of the 5' flanking region of the mouse serglycin gene on its ability to direct human growth hormone (hGH) expression in transfected cells. The solid, bold horizontal lines (————) represent the various lengths of the 5' flanking region of the mouse serglycin gene ligated to pφGH, a plasmid that contains a promoterless hGH gene (□□□□□). The negative and positive numbers in the various constructs refer to the length of the nucleotide sequence that extends upstream and downstream, respectively, of the transcription-initiation site of the gene. In each experiment, the amount of hGH was quantitated 4 d after transfection of rat basophilic leukemia-1 (RBL-1) cells and rat-1 fibroblasts with the specific plasmid construct. The numbers on the right are the hGH values obtained relative to the same population of cells transfected with the control plasmid construct, pXGH5. The indicated hGH activities represent the mean ± SD of data from 6 to 18 experiments. ND, not determined.

**B.** Blot analysis of hGH mRNA in rat basophilic leukemia-1 cells and rat-1 fibroblasts transfected with different DNA constructs. RNA blots containing approximately equal amounts of total RNA (10 μg/sample) from rat basophilic leukemia-1 cells and rat-1 fibroblasts transfected with pPG(-504/+24)hGH (lane 1), pPG(-118/+24)hGH (lane 2), pPG(-40/+24)hGH (lane 3), pφGH (lane 4), pSV40-hGH (lane 5), or pXGH5 (lane 6) were probed with a <sup>32</sup>P-labeled hGH cDNA. The arrows on the right indicate the migration positions of 2.0 kb rRNA, hGH mRNA, and β-actin mRNA. pSV40-hGH was obtained by Dr. J. Sarid.

-11-

Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

**Figure 9.** Effect of two 5' flanking regions of the mouse serglycin gene on its ability to enhance and suppress hGH production in cells transfected with plasmid constructs that contain an enhancerless SV40 early promoter. The hatched lines (□□□□□) and the round-dot lines (⊙⊙⊙⊙) represent the structural sequences of the hGH gene and SV40 early promoter, respectively, within the plasmid construct. The solid, bold horizontal lines (————) represent the specific parts of the 5' flanking-region of the mouse serglycin gene that is inserted upstream of the SV40 promoter in pSV40-hGH. The numbers on the right are the hGH values obtained at 4 d relative to those cells transfected with the control plasmid, pSV40-hGH. The indicated hGH activities represent the mean  $\pm$  SD values of data from 5 to 6 experiments of 4-d duration, with each experiment performed on 2-3 replicate dishes of cells.

**Figure 10.** Detection of *trans*-acting factors in the nucleus of rat basophilic leukemia-1 cells (RBL-1) and rat-1 fibroblasts (Rat-1 Fib.) that bind *cis*-acting elements in the putative suppressor region of the 5' flanking region of the mouse serglycin gene. Gel mobility shift assays were performed with the diagrammatically depicted nucleotide sequence in the 5' flanking region of the mouse serglycin gene. In lane 1, 1 ng of the <sup>32</sup>P-labeled DNA fragment (residues -250 to -161) was electrophoresed in the gel in the absence of nuclear extracts. In lanes 2 to 4 and lanes 5 to 7, the probe was incubated before electrophoresis with nuclear extracts from rat basophilic leukemia-1 cells and rat-1 fibroblasts, respectively. Competition assays were performed using 5 ng of the same nonradioactive DNA probe (lanes 3 and 6) or 100 ng of sonicated salmon sperm DNA (lanes 4 and 7). The probe and the *trans*-acting factors present in fibroblasts (F<sub>(-250/-161)</sub>-II and B/F<sub>(-250/-161)</sub>-I) and rat basophilic leukemia-1 cells (B/F<sub>(-250/-161)</sub>-I) are indicated on the right.

-12-

**Figure 11.** Detection of *trans*-acting factors in the nucleus of rat basophilic leukemia-1 cells and rat-1 fibroblasts that bind *cis*-acting elements in the putative enhancer region of the 5' flanking region of the mouse serglycin gene. Gel mobility shift assays were performed with the diagrammatically depicted nucleotide sequence in the 5' flanking region of the mouse serglycin gene. In lane 1, 1 ng of the <sup>32</sup>P-labeled DNA fragment (residues -118 to -81) was electrophoresed in the gel in the absence of nuclear extracts. In lanes 2 to 4 and lanes 5 to 7, the probe was incubated before electrophoresis with nuclear extracts from rat basophilic leukemia-1 cells and rat-1 fibroblasts, respectively. Competition assays were performed using 5 ng of the same nonradioactive DNA probe (lanes 3 and 6) or 100 ng of sonicated salmon sperm DNA (lanes 4 and 7). The probe, nonspecific (ns) bound probe, and the *trans*-acting factors present in fibroblasts (F<sub>(-118/-81)</sub>-I) and rat basophilic leukemia-1 cells (B<sub>(-118/-81)</sub>-I) are indicated on the right.

**Figure 12.** Role of residues -28, -30, and -38 in the proximal promoter region of the mouse serglycin gene in its interaction with *trans*-acting factors in the nuclei of rat basophilic leukemia-1 cells and rat-1 fibroblasts. Gel mobility shift assays were performed with the diagrammatically depicted 64 bp nucleotide sequence in the 5' flanking region of the mouse serglycin gene (residues -40 to +24) prepared with and without point mutations. In lane 1, 1 ng of the <sup>32</sup>P-labeled oligonucleotide was electrophoresed in the gel in the absence of nuclear extracts. In lanes 2 to 6 and lanes 7 to 11, the probe was incubated before electrophoresis with nuclear extracts from rat basophilic leukemia-1 cells and rat-1 fibroblasts, respectively. Competition assays were performed with 5 ng of nonradioactive DNA that corresponded to the probe (lanes 3 and 8) or 50 ng of nonradioactive DNA that had a mutated residue -30 (lanes 4 and 9), -28 (lanes 5 and 10), or -38 (lanes 6 and 11). The probe and the retarded *trans*-acting factors present in fibroblasts

-13-

(F<sub>(-40/+24)</sub>-II and B/F<sub>(-24/+24)</sub>-I) and rat basophilic leukemia-1 cells (B<sub>(-40/+24)</sub>-II and B/F<sub>(-40/+24)</sub>-I) are indicated on the right.

**Figure 13.** Location of the *Alu* Elements and the *HpaII/MspI* Sites in the Human Serglycin Gene. Figure 13A: the locations of the 21 *Alu* elements in the 5'-flanking region, intron 1, and intron 2 of the serglycin gene are depicted. The locations of the two *Alu* elements containing only the left arm are identified with 1/2. The three exons are boxed (■). Figure 13B: the locations of the *HpaII/MspI* sites (5'-CCGG-3') in the serglycin gene are indicated by the vertical lines. The letters depict the location of the probes used to determine the extent of methylation of these sites. Sites in the serglycin gene in HL-60 cells that are at least partially methylated are indicated by closed circles, and nonmethylated sites are indicated by open circles.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

### I. Definitions

In the description that follows, a number of terms used in recombinant DNA (rDNA) technology are extensively utilized. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

**Gene.** A DNA sequence containing a template for a RNA polymerase. The RNA transcribed from a gene may or may not code for a protein. RNA that codes for a protein is termed messenger RNA (mRNA) and, in eukaryotes, is transcribed by RNA polymerase II. However, it is also known to construct a gene containing a RNA polymerase II template wherein a RNA sequence is transcribed which has a sequence complementary to that of a specific mRNA but is not normally

-14-

translated. Such a gene construct is herein termed an "antisense RNA gene" and such a RNA transcript is termed an "antisense RNA." Antisense RNAs are not normally translatable due to the presence of translational stop codons in the antisense RNA sequence.

5           A "complementary DNA" or "cDNA" gene includes recombinant genes synthesized by reverse transcription of mRNA and from which intervening sequences (introns) have been removed.

Cloning vehicle. A plasmid or phage DNA or other DNA sequence which is able to replicate autonomously in a host cell, and that  
10           is characterized by one or a small number of endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion without loss of an essential biological function of the vehicle, and into which DNA may be spliced in order to bring about its replication and cloning. The cloning vehicle may further contain a marker suitable for  
15           use in the identification of cells transformed with the cloning vehicle. Markers, for example, are tetracycline resistance or ampicillin resistance. The word "vector" is sometimes used for "cloning vehicle."

Expression vehicle. A vehicle or vector similar to a cloning vehicle but which is capable of expressing a gene that has been cloned into it,  
20           after transformation into a host. Accordingly to the invention, the cloned gene or coding sequence (the gene of interest) is usually placed under the control of (i.e., operably linked to) certain control sequences such as the promoter sequences and/or regulatory elements of the invention. Expression control sequences will vary and may additionally contain  
25           transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements in addition to those of the invention, and/or translational initiation and termination sites.

Proteoglycan. This term as used throughout the specification and claims means mammalian and especially human "hematopoietic cell  
30           proteoglycan" that contains glycosaminoglycan chains covalently bound to the proteoglycan's core protein.



-15-

Serglycin. Serglycin is the peptide core of hematopoietic cell secretory granule proteoglycan. The term is meant to include peptide fragments of hematopoietic cell secretory granule proteoglycan wherein the peptide core protein contains less than the naturally-occurring number of amino acids, but which retains biological (functional or structural) activity. Example of the functional activity serglycin is the ability to induce a specific biological response in the same manner that the native non-recombinant protein does, such as the ability to be conjugated into a specific proteoglycan form. An example of a structural activity is the ability to bind antibodies which also recognize the native non-recombinant protein.

The term is also used to include serglycin fusion proteins, that is, a peptide which comprises the sequence of a naturally-occurring serglycin or a biologically active fragment thereof together with one or more additional flanking amino acids, but which still possesses hematopoietic cell secretory granule proteoglycan biological (functional or structural) activity.

Transcriptional regulatory element. A transcriptional regulatory element (or DNA regulatory element) is a DNA sequence that, when operably linked to a gene of interest, is capable of altering the transcription of such gene of interest in a specific way characteristic of such element. Transcriptional regulatory elements include promoters, enhancers, suppressors, transcriptional start sites, transcriptional stop sites, polyadenylation sites, and the like.

Functional Derivative. A "functional derivative" of the DNA regulatory elements of the invention is a DNA sequence that possesses at least a biologically active fragment of the sequence of the regulatory elements of the invention; by "biologically active" fragment is meant that the fragment retain a biological activity (either functional or structural) that is substantially similar to a biological activity of the full-length DNA element. A biological activity of a DNA regulatory element of the

-16-

invention is its ability to alter transcription in a manner known to be attributed to the full-length element. Hence, biologically active fragments of the suppressor element of the invention will retain the ability to inhibit or repress transcription; biologically active fragments of the enhancer  
5 element of the invention will retain the ability to stimulate transcription; and biologically active fragments of the promoter sequence of the invention will retain the ability to promote transcription.

The term "functional derivative" is intended to include the "fragments," "variants," "analogues," or "chemical derivatives" of a molecule.

10 Fragment. A "fragment" of a nucleotide or peptide sequence is meant to refer to a sequence that is less than that believed to be the "full-length" sequence.

Variant. A "variant" of a molecule is meant to refer to allelic variations of such sequences, that is, a sequence substantially similar in  
15 structure and biological activity to either the entire molecule, or to a fragment thereof.

## II. Genetic Engineering of the Regulatory Elements of the Invention

Provided herein are transcriptional *cis*-acting elements of hematopoietic cells: an enhancer element, a suppressor element and a  
20 novel promoter element. The transcriptional *cis*-acting elements of the invention are naturally found in the 5' regulatory region of the serglycin gene. In addition, provided herein are *trans*-acting factors, such factors specifically binding to the *cis*-acting elements of the invention. The process for genetically engineering the genetic regulatory elements of the  
25 invention, or the *trans*-acting factors of the invention, is facilitated through the cloning of genetic sequences that contain the sequence for such regulatory elements or factors. The 333 base pair (bp) nucleotide sequence 5' of the transcription-initiation site of the mouse gene is nearly identical to the corresponding region of the human gene (Nicodemus

*et al.*, *J. Biol. Chem.* 265:5889-5896 (1990)). Thus, the mouse or human sequence may be used interchangeably for the cloning of the alternate species, or the cloning of similar sequences in other species.

5 The regulatory elements may be cloned directly, using, for example, promoter probe vectors and the like. For the identification of those regions of the serglycin gene that provide the *cis*-acting enhancer, suppressor, and/or promoter, rat basophilic leukemia-1 cells, mouse WEHI-3 cells, rat-1 fibroblasts, and mouse 3T3 fibroblasts may be transiently transfected with plasmid constructs containing various lengths  
10 of the 504-bp 5' flanking region of the mouse serglycin gene linked to a gene of interest, for example, the human growth hormone (hGH) gene, so as to provide a reporter, expression of which indicates transcriptional activity. Rat basophilic leukemia-1 cells and mouse WEHI-3 cells are preferred because they contain cytoplasmic granules and express large  
15 amounts of serglycin, whereas no serglycin transcript is present in either fibroblast line (Tantravahi *et al.*, *Proc. Natl. Acad. Sci. USA* 83:9207-9210 (1986)).

The hGH transient expression system is preferred because it is at least 10-fold more sensitive than the CAT system (Selden *et al.*, *Mol. Cell Biol.* 6:3173-3179 (1986)) or other systems that are based on the  
20 expression of  $\beta$ -galactosidase (An *et al.*, *Mol. Cell. Biol.* 2:1628-1632 (1982)) and xanthine-guanine phosphoribosyl transferase (Chu *et al.*, *Nucleic Acids Res.* 13:2921-2930 (1985)). This increased sensitivity enables hGH levels to be measured after transfection with a very small amount of  
25 plasmid, thus avoiding potential problems of competition (Selden *et al.*, *Mol. Cell Biol.* 6:3173-3179 (1986)). The hGH transient expression system is well-suited for use because the plasmids are known in the art (such as pXGH5 for example) that can be used as an internal positive control for normalizing the efficiency of transfection. thereby facilitating the  
30 interpretation of data from separate experiments.

-18-

The specificity of the transcriptional regulatory elements of the invention is such that reporter mRNA expression using the transcriptional regulatory elements of the invention may be detected in rat-1 fibroblasts transfected with appropriate promoter probe plasmids that contain the desired regulatory elements operably linked to the reporter gene. For example, construct pPG(-118/+24)hGH, containing the promoter and enhancer element of the invention will provide high levels of reporter expression in such host cells.

In addition, high levels of reporter expression will be detected in rat basophilic leukemia-1 cells transfected with constructs containing the suppressor, enhancer, and promoter of the invention, (for example, pPG(-504/+24)hGH) or just the enhancer and promoter of the invention (for example pPG(-118/+24)hGH).

In contrast, lesser amounts reporter mRNA will be detected in rat basophilic leukemia-1 cells and rat-1 fibroblasts transfected with promoter probe vectors containing only the promoter of the invention (for example, pPG(-40/+24)hGH). No reporter mRNA will be detected in rat basophilic leukemia-1 cells transfected with p $\phi$ GH or in rat-1 fibroblasts transfected with constructs containing the suppressor, enhancer and promoter of the invention (for example pPG(-504/+24)hGH) or p $\phi$ GH. Because large amounts of hGH are detected in the culture media of rat basophilic leukemia-1 cells and rat-1 fibroblasts that contain abundant levels of hGH mRNA and because lesser amounts of hGH are detected in the culture media of cells containing intermediate levels of hGH mRNA, transcription and translation of the hGH gene are related in both transfected cell types.

The results of the transfections should be normalized to that obtained with a reference plasmid, such as, for the growth hormone reporter, pXGH5. Rat basophilic leukemia-1 cells produce more reporter (18-fold more hGH) than transfected rat-1 fibroblasts. Likewise, mouse WEHI-3 cells produce more reporter than transfected mouse 3T3

fibroblasts (pPG(-504/+24)hGH produced 20-fold more hGH in mouse WEHI-3 cells than transfected mouse 3T3 fibroblasts).

Based on the results such as those discussed above and herein, the presence of *cis*-acting regulatory elements the 5' flanking region of a gene, and/or in the first intron of such gene, and especially of the serglycin gene, may be established. Such serglycin gene elements preferentially enhance the constitutive transcription of a gene of interest in hematopoietic cells or preferentially suppress transcription of a gene of interest in fibroblasts, although small differences may be present, due to other factors.

The sequences of intron 1 of the serglycin gene may act in concert to regulate transcription of a gene that is operably linked to the serglycin gene promoter in different cell types. When it is desired to utilize this concerted action, intron 1 sequences of the serglycin gene may be inserted into the coding sequence of a gene of interest such that what becomes exon 1 has approximately the same size as exon 1 of serglycin, and in a manner such that the reading frame of the coding sequence is not altered, and the normal recognition sequences at the flanking regions of the intron are provided, so as to allow subsequent excision of the intron.

To locate the *cis*-acting elements of the invention more precisely, additional plasmid constructs may be prepared that contain progressively less of the 5' flanking region of the serglycin gene. For example, the transfection of rat basophilic leukemia-1 cells and rat-1 fibroblasts with these shortened constructs of the 5' flanking region of the mouse serglycin gene reveal that a *cis*-acting element resides between residues -250 and -190 and suppresses transcription of this gene, and that this suppressor element is more dominantly active in rat-1 fibroblasts than in rat basophilic leukemia-1 cells. In a similar manner, such experiments revealed that an enhancer element resides between residues -118 and -81 of the mouse serglycin gene that not only appears to be important for the

-20-

positive constitutive transcription of this gene but also is dominantly active in rat basophilic leukemia-1 cells.

5 Rat basophilic leukemia-1 cells and fibroblasts produce substantially more reporter protein when transfected with a construct containing the enhancer and promoter of the invention (for example, pPG(-118/44)-SV40-hGH) than with a control that provides a foreign  
10 element (for example pSV40-hGH). Typical of other enhancers, the enhancer activity of the enhancer of the invention is not diminished by changing its orientation and its distance from the SV40 early promoter in the plasmid.

The promoter element of the invention is a unique sequence that provides an alternate to the classical TATA box. For most genes, transcription is initiated -30 bp downstream of the proximal end of the promoter, which usually is a TATA box. Because no reporter protein is  
15 detected when rat basophilic leukemia-1 cells and rat-1 fibroblasts are transfected with constructs containing only 20 nucleotides of the proximal end of the serglycin gene, (pPG(-20/+24)hGH), but some reporter protein is produced by cells transfected with constructs containing at least 40 nucleotides (pPG(-40/+24)hGH), the proximal element of the promoter  
20 of the invention resides between residues -40 and -20. Inasmuch as no TATA box is present in this region (Avraham *et al.*, *J. Biol. Chem.* 264:16719-16726 (1989); Nicodemus *et al.*, *J. Biol. Chem.* 265:5889-5896 (1990)), the TCTAAAA sequence at residues -31 to -25 may serve as an alternative element.

25 To demonstrate the important residues of the elements of the invention, they may be mutated, using techniques known in the art. For example, to demonstrate the important residues of the promoter element of the invention, residues -28, -30, or -38 were mutated. Based on the relative amount of reporter produced in the transfected cells, it  
30 was shown that the 5' flanking region containing the TCTAAAA sequence functions as a TATA box equivalent.

Alternatively to using promoter probe vectors for the cloning of the regulatory elements of the invention, the coding sequence of the serglycin gene (previously called the secretory granule proteoglycan peptide core protein) may be cloned and used to identify clones or DNA containing  
5 the desired regulatory elements operably linked to the cloned coding sequence. The discussion below, while it specifically refers to cloning of the serglycin gene, may also be adapted by those of skill in the art for the cloning of the *trans*-acting factors of the invention.

The regulatory elements of the genomic DNA of the invention may  
10 be obtained in association with the 5' promoter region of the serglycin gene. For example, when rat-1 fibroblasts were stably transfected with the mouse genomic clone,  $\lambda$ -MG-PG1, two cell lines were obtained that expressed low levels of the 1.0-kb serglycin mRNA (Avraham *et al.*, *J. Biol. Chem.* 264:16719-16726 (1989)). This finding indicated that  
15  $\lambda$ -MG-PG1 contained the entire mouse serglycin gene, including, perhaps, some of the regulatory elements within its promoter region. S1 nuclease mapping and primer extension analysis revealed that the primary transcription-initiation site for this gene in mouse bone marrow-derived mast cells (BMC) resides -40 nucleotides upstream of the translation-  
20 initiation site (Avraham *et al.*, *J. Biol. Chem.* 264:16719-16726 (1989)).

As used herein, the term "genetic sequences" is intended to refer to a nucleic acid molecule (preferably DNA). Genetic sequences that are capable of providing the regulatory elements of the invention are derived from a variety of sources, including genomic DNA, synthetic DNA, and  
25 combinations thereof. Genetic sequences that are capable of encoding serglycin may further be derived from mRNA or cDNA.

Genomic DNA containing the serglycin gene can be extracted and purified from any eukaryotic and especially mammalian cell that has this gene in its genome by means well known in the art (for example, see  
30 *Guide to Molecular Cloning Techniques*. S.L. Berger *et al.*, eds., Academic Press (1987)).

-22-

Serglycin mRNA can be isolated from any cell which produces or expresses this protein. Serglycin mRNA can also be used to produce cDNA by means well known in the art (for example, see *Guide to Molecular Cloning Techniques*, S.L. Berger *et al.*, eds., Academic Press (1987)). Such cell sources include, but are not limited to, fresh cell preparations and cultured cell lines, especially fresh or cultured connective tissue mast cells, mucosal mast cells, basophils, natural killer cells, cytotoxic T lymphocytes, eosinophils, neutrophils, macrophages and platelets.

Preferably, the mRNA preparation used will be enriched in mRNA coding for serglycin, either naturally, by isolation from a cells which are producing large amounts of the protein, or *in vitro*, by techniques commonly used to enrich mRNA preparations for specific sequences, such as sucrose gradient centrifugation, or both.

For cloning into a vector, suitable DNA preparations (genomic DNA containing the regulatory elements of the invention or cDNA encoding the serglycin) are randomly sheared or enzymatically cleaved, respectively, and ligated into appropriate vectors to form a recombinant gene (either genomic or cDNA) library.

A DNA sequence providing the regulatory elements of the invention, or the serglycin coding sequence, may be inserted into a DNA vector in accordance with conventional techniques, including blunt-ending or staggered-ending termini for ligation, restriction enzyme digestion to provide appropriate termini, filling in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and ligation with appropriate ligases. Techniques for such manipulations are disclosed by Maniatis, T., *et al.*, *supra*, and are well known in the art.

A serglycin clone may be identified by any means which specifically selects for serglycin DNA such as, for example, a) by hybridization with an appropriate nucleic acid probe(s) containing a sequence specific for the DNA of this protein, or b) by hybridization-selected translational analysis



-23-

in which native mRNA which hybridizes to the clone in question is translated *in vitro* and the translation products are further characterized, or, c) if the cloned genetic sequences are themselves capable of expressing mRNA, by immunoprecipitation of a translated serglycin protein product produced by the host containing the clone. The ability to specifically bind antibody against serglycin or its proteoglycan, the ability to elicit the production of antibody capable of binding to serglycin as its proteoglycan, and/or the ability to provide a serglycin proteoglycan-associated function to a recipient cell, are all examples of the biological properties of the serglycin proteoglycan.

Oligonucleotide probes specific for the serglycin gene are useful for the identification of genomic or cDNA clones to this protein, or useful for the identification of clones to the regulatory elements of the invention, can be designed from knowledge of the amino acid sequence of the protein's peptide core, or from knowledge of the nucleotide sequence of the regulatory element, respectfully.

When designing a probe against a peptide sequence, the sequence of amino acid residues in a peptide is designated herein either through the use of their commonly employed three-letter designations or by their single-letter designations. A listing of these three-letter and one-letter designations may be found in textbooks such as *Biochemistry*, Lehninger, A., Worth Publishers, New York, NY (1970). When the amino acid sequence is listed horizontally, the amino terminus is intended to be on the left end whereas the carboxy terminus is intended to be at the right end. The residues of amino acids in a peptide may be separated by hyphens. Such hyphens are intended solely to facilitate the presentation of a sequence.

When designing probes against a peptide sequence, because the genetic code is degenerate, more than one codon may be used to encode a particular amino acid (Watson, J.D., In: *Molecular Biology of the Gene*, 3rd Ed., W.A. Benjamin, Inc., Menlo Park, CA (1977), pp. 356-357). The

peptide fragments are analyzed to identify sequences of amino acids that may be encoded by oligonucleotides having the lowest degree of degeneracy. This is preferably accomplished by identifying sequences that contain amino acids which are encoded by only a single codon.

5           Although occasionally an amino acid sequence may be encoded by only a single oligonucleotide sequence, frequently the amino acid sequence may be encoded by any of a set of similar oligonucleotides. Importantly, whereas all of the members of this set contain oligonucleotide sequences that are capable of encoding the same peptide  
10           fragment and, thus, potentially contain the same oligonucleotide sequence as the gene which encodes the peptide fragment, only one member of the set contains the nucleotide sequence that is identical to the exon coding sequence of the gene. Because this member is present within the set, and is capable of hybridizing to DNA even in the presence of the other  
15           members of the set, it is possible to employ the unfractionated set of oligonucleotides in the same manner in which one would employ a single oligonucleotide to clone the gene that encodes the peptide.

          Using the genetic code (Watson, J.D., In: *Molecular Biology of the Gene*, 3rd Ed., W.A. Benjamin, Inc., Menlo Park, CA (1977)), one or  
20           more different polynucleotides or oligonucleotides can be identified from the amino acid sequence, each of which would be capable of encoding the serglycin protein or fragments thereof. The probability that a particular polynucleotide will, in fact, constitute the actual serglycin encoding sequence can be estimated by considering abnormal base pairing  
25           relationships and the frequency with which a particular codon is actually used (to encode a particular amino acid) in eukaryotic cells. Such "codon usage rules" are disclosed by Lathe, R., *et al*, *J. Molec. Biol.* 183:1-12 (1985). Using the "codon usage rules" of Lathe, a single polynucleotide sequence, or a set of polynucleotide sequences, that contain a theoretical  
30           "most probable" nucleotide sequence capable of encoding the serglycin sequences is identified.

The suitable polynucleotide, or set of polynucleotides, that is capable of encoding the serglycin gene, or fragment thereof, may be synthesized by means well known in the art (see, for example, *Synthesis and Application of DNA and RNA*, S.A. Narang, ed., 1987, Academic Press, San Diego, CA) and employed as a probe to identify and isolate the cloned serglycin gene by techniques known in the art. Techniques of nucleic acid hybridization and clone identification are disclosed by Maniatis, T., *et al.*, (In: *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY (1982)), and by Hames, B.D., *et al.*, (In: *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, DC (1985)), which references are herein incorporated by reference. Those members of the above-described gene library that are found to be capable of such hybridization are then analyzed to determine the extent and nature of the serglycin encoding sequences that they contain.

To facilitate the detection of the desired serglycin DNA encoding sequence, the above-described DNA probe may be labeled with a detectable group. Such detectable group can be any material having a detectable physical or chemical property. Such materials have been well-developed in the field of nucleic acid hybridization and in general most any label useful in such methods can be applied to the present invention. Particularly useful are radioactive labels, such as  $^{32}\text{P}$ ,  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{35}\text{S}$ ,  $^{125}\text{I}$ , or the like. Any radioactive label may be employed which provides for an adequate signal and has a sufficient half-life. The oligonucleotide may be radioactively labeled, for example, by "nick-translation" by well-known means, as described in, for example, Rigby, P.J.W., *et al.*, *J. Mol. Biol.* 113:237 (1977) and by T4 DNA polymerase replacement synthesis as described in, for example, Deen, K.C., *et al.*, *Anal. Biochem.* 135:456 (1983).

Alternatively, polynucleotides are also useful as nucleic acid hybridization probes when labeled with a non-radioactive marker such as

biotin, an enzyme or a fluorescent group. See, for example, Leary, J.J., *et al.*, *Proc. Natl. Acad. Sci. USA* 80:4045 (1983); Renz, M., *et al.*, *Nucl. Acids Res.* 12:3435 (1984); and Renz, M., *EMBO J.* 6:817 (1983).

Thus, in summary, the actual identification of the amino acid  
5 sequence of serglycin permits the identification of a theoretical "most  
probable" DNA sequence, or a set of such sequences, capable of encoding  
such a peptide. By constructing an oligonucleotide complementary to this  
theoretical sequence (or by constructing a set of oligonucleotides  
complementary to the set of "most probable" oligonucleotides), one  
10 obtains a DNA molecule (or set of DNA molecules), capable of function-  
ing as a probe(s) for the identification and isolation of clones containing  
the serglycin gene, and thus the regulatory elements of the invention.

In an alternative way of cloning the serglycin gene, a library is  
prepared using an expression vector, by cloning DNA prepared from a cell  
15 possessing, and preferably, capable of expressing, serglycin, into an  
expression vector. The cDNA library is then screened for members that  
express serglycin, for example, by screening the library with antibodies to  
the protein, such as the antibody depicted in Figure 6.

In another embodiment, a previously described rat L2 cell-derived  
20 cDNA of a related proteoglycan, pPG-1, (disclosed in Bourdon *et al.*,  
*Proc. Natl. Acad. Sci. USA* 82:1322 (1985)) is used to identify a sequence  
encoding serglycin. For example, Southern blots of digested genomic  
DNA may be probed with nick-translated pPG-1 or pPG-M (a gene  
specific 489 bp *Ssp* I → 3' end fragment of pPG-1), Tantravahi *et al.*,  
25 *Proc. Natl. Acad. Sci. USA* 83:9207 (1986), under reduced stringency if  
necessary, to allow for mismatch between the sequence expressed in the  
different species.

The above discussed methods are, therefore, capable of identifying  
genetic sequences that are capable of encoding the serglycin or fragments  
30 of this protein. Such coding sequences may then be used to identify  
clones containing the transcriptional regulatory elements of the invention.

-27-

For example, using the techniques described above, a number of DNA fragments were identified when a human genomic DNA blot was probed under conditions of low stringency with a rat serglycin cDNA (Stevens *et al.*, *J. Biol. Chem.* 263:7287 (1988)). Nevertheless, by probing a human promyelomonocytic HL-60 cell-derived cDNA library under conditions of low stringency with the rat cDNA, a cDNA was isolated and characterized that encodes human serglycin (Stevens *et al.*, *J. Biol. Chem.* 263:7287 (1988)). Sequence analysis of a resulting cDNA clone indicated that in the human this proteoglycan peptide core is only 17.6 kDa and contains an 18 amino acid glycosaminoglycan attachment region consisting primarily of alternating serine and glycine. A single gene that resides on chromosome 10 encodes this human protein (Stevens *et al.*, *J. Biol. Chem.* 263:7287 (1988); Nicodemus *et al.*, *J. Biol. Chem.* 265:5889 (1990); Mattei *et al.*, *Human Genetics* 82:87 (1989)). A human genomic library was probed under conditions of high stringency with a 5' fragment of the HL-60 cell cDNA to isolate two 18-kb genomic fragments that taken together contain the entire human serglycin gene (Nicodemus *et al.*, *J. Biol. Chem.* 265:5889 (1990)). A restriction map of this human gene was constructed, and the genomic fragments subcloned into Bluescript™ plasmid, and the nucleotide sequence of the entire 16.6 kb human gene determined plus 0.7 kb of 5' flanking DNA, using techniques known in the art.

In addition, a 1.0-kb cDNA that encodes mouse serglycin was isolated from a mouse bone marrow-derived mast cells-derived cDNA library. When the predicted amino acid sequences of the mouse, rat, and human serglycin were compared, the N-terminus (not the serine-glycine rich glycosaminoglycan-attachment region) was found to be the most conserved region. This surprising finding suggests that N terminus of the translated peptide core is important for the structure, function, and/or metabolism of this family of proteoglycans. Areas of identity in the 3' and 5' untranslated regions in the human, rat, and mouse proteoglycan cDNAs were also observed (Avraham *et al.*, *Proc. Natl. Acad. Sci. USA* 86:3763

(1989)). Interestingly, these 3' and 5' conserved untranslated nucleotide sequences were almost identical in the corresponding regions of the cDNAs that encode human mast cell tryptase (Miller *et al.*, *J. Clin. Invest.* 84:1188 (1989)), dog mast cell tryptase (Vanderslice *et al.*, *Biochemistry* 28:4148 (1989)), mouse mast cell protease-2 (Serafin *et al.*, *J. Biol. Chem.* 265:423 (1990)), and rat mast cell protease-II (Benfey *et al.*, *J. Biol. Chem.* 262:5377 (1987)), suggesting that these nucleotide sequences may be important for coordinated regulation of those genes that encode proteins destined to reside in the secretory granules of hematopoietic cells.

To isolate the mouse serglycin gene, a mouse genomic DNA library was probed under conditions of high stringency with a 3' gene specific fragment of the mouse bone marrow-derived mast cells-derived cDNA. An ~18 kb genomic clone (kMG-PG1) which contains the entire gene that encodes mouse serglycin was isolated. The exon/intron organization of the mouse gene was determined, as well as the transcription-initiation site and the 504-bp nucleotide sequence that is upstream of the gene.

Typically, transcription is initiated ~30 bp downstream of an element within the proximal end of a gene's promoter (defined in this case as the smallest amount of nucleotide sequence that must be present to get minimal transcription of a gene in a cell). In most eukaryotic genes, their promoters contain either a TATA box (Breathnach *et al.*, *Ann. Rev. Biochem.* 50:349 (1981)) or a GC-rich element (Sehgal *et al.*, *Mol. Cell. Biol.* 8:3160 (1988)). In rarer cases, such as the terminal deoxynucleotidyltransferase gene (Smale, S.T., and Baltimore, D. (1989) *Cell* 57:103 (1988)), a third type of promoter region is present which lacks these specific transcription-initiation control sequences.

The mouse serglycin gene does not contain either a classical TATA box or a GC-rich element ~30 bp upstream of the transcription-initiation site. Therefore, its promoter appears to belong to the rarer, third class of promoters.

-29-

Accordingly, the above discussed methods are also capable of directly identifying clones containing genetic sequences containing the transcriptional regulatory elements of the invention.

5 The above discussed methods are also capable of being adapted for the identification of clones directed to the *trans*-acting factors of the invention. Especially, clones capable of expressing such *trans*-acting factors may be identified utilizing the target sequence to which they bind (in a double-stranded DNA form) to detect their presence in protein-DNA binding assays. Such assays are well known in the art.

10 In order to further characterize such genetic sequences, and, in order to produce recombinant protein under the transcriptional control of such sequences, such transcriptional regulatory elements must be provided to an appropriate host.

15 III. Expression of Proteins Operably-linked to the Transcriptional Regulatory Elements of the Invention

As used herein, "heterologous protein" is intended to refer to a peptide sequence that is heterologous to the transcriptional regulatory elements of the invention. A skilled artisan will recognize that, if desired, the teaching herein will also apply to the expression of genetic sequences encoding serglycin homologous to such regulatory elements.

20 To express a heterologous protein under the control of the transcriptional regulatory elements of the invention, the heterologous protein must be "operably-linked" to the regulatory element. An operable linkage is a linkage in which a desired sequence is connected to a transcriptional or translational regulatory sequence (or sequences) in such a way as to place expression (or operation) of the desired sequence under the influence or control of the regulatory sequence.

25 Two DNA sequences (such as a sequence encoding a heterologous protein and a promoter region sequence linked to the 5' end of the

-30-

encoding sequence) are said to be operably linked if induction of promoter function results in the transcription of the DNA encoding the heterologous protein and if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the expression regulatory sequences to direct the expression of the heterologous protein DNA, or (3) interfere with the ability of the heterologous protein template to be transcribed by the promoter region sequence. Thus, a promoter region would be operably linked to a DNA sequence if the promoter were capable of effecting transcription of that DNA sequence.

In a similar manner, a transcriptional regulatory element that stimulated or repressed promoter function may be operably-linked to such promoter. Exact placement of the element in the nucleotide chain is not critical as long as the element is located at a position from which the desired effects on the operably linked promoter may be revealed. A nucleic acid molecule, such as DNA, is said to be "capable of expressing" a polypeptide if it contains expression control sequences which contain transcriptional regulatory information and such sequences are operably linked to the nucleotide sequence which encodes the polypeptide.

For the complete control of heterologous gene expression, all transcriptional and translational regulatory elements (or signals) that are operably linked to a heterologous gene should be recognizable by the appropriate host. By "recognizable" in a host is meant that such signals are functional in such host.

The cloned transcriptional regulatory elements, obtained through the methods described above, and preferably in a double-stranded form, may be operably linked to a heterologous gene, preferably in an expression vector, and introduced into a host cell, preferably eukaryote cell, and most preferably, a eukaryotic cell of the hematopoietic cell origin, to produce recombinant heterologous protein.



Expression of the heterologous protein in different hosts may result in different post-translational modifications that may or may not alter the properties of the heterologous protein. Especially preferred hosts are cells either *in vivo*, or in tissue culture that provide post-translational  
5 modifications to the heterologous protein that include folding and/or glycosylation at sites similar or identical to that found for the native proteoglycan.

Appropriate cells of hematopoietic cell origin include, for example, hematopoietic cells that participate in immune and inflammatory  
10 responses, including connective tissue mast cells, mucosal mast cells, basophils, natural killer cells, cytotoxic T lymphocytes, eosinophils, neutrophils, macrophages, and platelets. For example, rat basophilic leukemia-1 cells (ATCC CRL-1378), mouse bone marrow derived mast cells, mouse mast cells immortalized with Kirsten sarcoma virus, normal  
15 mouse mast cells that have been co-cultured with mouse fibroblasts, or mouse myelomonocytic WEHI-3 cells (ATCC TIB-68) are useful. Razin *et al.*, *J. Immun.* 132:1479 (1984); Levi-Schaffer *et al.*, *Proc. Natl. Acad. Sci. (USA)* 83:6485 (1986) and Reynolds *et al.*, "Immortalization of Murine Connective Tissue-type Mast Cells at Multiple Stages of Their  
20 Differentiation by Coculture of Splenocytes with Fibroblasts that Produce Kirsten Sarcoma Virus," *J. Biol. Chem.* 263:12783-12791 (1988). See Example 5, below. Methods for the long term *in vitro* proliferation of pluripotent bone marrow stem cells are known (*Handbook of the Hematopoietic Microenvironment*, M. Tavassoli, ed., Humana Press, Inc.,  
25 Clifton, New Jersey, 1989).

The precise nature of the regulatory regions needed for gene expression may vary between species or cell types, but shall in general include, as necessary, 5' non-transcribing and 5' non-translating (non-coding) sequences involved with initiation of transcription and translation  
30 respectively. Especially, at a minimum, such 5' non-transcribing control

-32-

sequences will include a region which contains a promoter for transcriptional control of the operably linked gene.

5 The promoter preferably is the serglycin gene promoter of the invention. However, the enhancer and suppressor transcriptional regulatory elements of the invention may be operably linked to any promoter that is function in the desired host cell. A wide variety of transcriptional and translational regulatory sequences can be employed, operably linked to a transcriptional regulatory element of the invention, depending upon the nature of the eukaryotic host. In eukaryotes, 10 where transcription is not linked to translation, such control regions may or may not provide an initiator methionine (AUG) codon, depending on whether the operably linked heterologous sequence contains such a methionine. Such regions will, in general, include a promoter region sufficient to direct the initiation of RNA synthesis in the host cell. 15 Promoters from heterologous mammalian genes that encode an mRNA product capable of translation are preferred, and especially, strong promoters such as the promoter for actin, collagen, myosin, etc., can be employed provided they also function as promoters in the host cell, and provided that their function is also capable of being control by the desired 20 positive or suppressor of the invention.

As is widely known, translation of eukaryotic mRNA is initiated at the codon that encodes the first methionine. For this reason, it is preferable to ensure that the linkage between a eukaryotic promoter and a DNA sequence that encodes the heterologous protein does not contain 25 any intervening codons that are capable of encoding a methionine. The presence of such codons results either in a formation of a fusion protein (if the AUG codon is in the same reading frame as the DNA encoding the heterologous protein) or a frame-shift mutation (if the AUG codon is not in the same reading frame as the DNA encoding the heterologous 30 protein.

-33-

If desired, a fusion product of the heterologous protein may be constructed. For example, the sequence coding for the heterologous protein may be linked to a signal sequence which will allow secretion of the protein from, or the compartmentalization of the protein in, a particular host. Such signal sequences may be designed with or without specific protease sites such that the signal peptide sequence is amenable to subsequent removal. Alternatively, the native signal sequence for this protein may be used.

The transcriptional initiation regulatory elements of the invention can be selected to allow for repression or activation, so that expression of the operably linked genes can be modulated. Translational signals are not necessary when it is desired to express antisense RNA sequences.

If desired, the non-transcribed and/or non-translated regions 3' to the sequence coding for the heterologous protein can be obtained by the above-described cloning methods. The 3'-non-transcribed region may be retained for its transcriptional termination regulatory sequence elements; the 3-non-translated region may be retained for its translational termination regulatory sequence elements, or for those elements that direct polyadenylation in eukaryotic cells. Where the native expression control sequences signals do not function satisfactorily host cell, then sequences functional in the host cell may be substituted.

To transform a mammalian cell with the DNA constructs of the invention many vector systems are available, depending upon whether it is desired to insert the heterologous protein DNA construct into the host cell chromosomal DNA, or to allow it to exist in an extrachromosomal form.

If the heterologous protein's DNA sequence and an operably linked promoter is introduced into a recipient eukaryotic cell as a non-replicating DNA (or RNA) molecule, which may either be a linear molecule or, more preferably, a closed covalent circular molecule that is incapable of autonomous replication, the expression of the heterologous

protein may occur through the transient expression of the introduced sequence.

Genetically stable transformants may be constructed with vector systems, or transformation systems, whereby the heterologous protein's DNA is integrated into the host chromosome. Such integration may occur  
5 *de novo* within the cell or, in a most preferred embodiment, be assisted by transformation with a vector that functionally inserts itself into the host chromosome. Vectors capable of chromosomal insertion include, for example, retroviral vectors, transposons or other DNA elements which  
10 promote integration of DNA sequences in chromosomes, especially DNA sequence homologous to a desired chromosomal insertion site.

Cells that have stably integrated the introduced DNA into their chromosomes are selected by also introducing one or more markers that allow for selection of host cells which that the desired sequence. For  
15 example, the marker may provide biocide resistance, e.g., resistance to antibiotics, or heavy metals, such as copper, or the like. The selectable marker gene can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection.

In another embodiment, the introduced sequence is incorporated  
20 into a plasmid or viral vector capable of autonomous replication in the recipient host. Any of a wide variety of vectors may be employed for this purpose, as outlined below.

Factors of importance in selecting a particular plasmid or viral vector include: the ease with which recipient cells that contain the vector  
25 may be recognized and selected from those recipient cells which do not contain the vector; the number of copies of the vector which are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

Preferred eukaryotic plasmids include those derived from the  
30 bovine papilloma virus, vaccinia virus, and SV40. Such plasmids are well known in the art and are commonly or commercially available. For

example, mammalian expression vector systems in which it is possible to cotransfect with a helper virus to amplify plasmid copy number, and, integrate the plasmid into the chromosomes of host cells have been described (Perkins, A.S. *et al.*, *Mol. Cell Biol.* 3:1123 (1983); Clontech, Palo Alto, California).

Once the vector or DNA sequence containing the construct(s) is prepared for expression, the DNA construct(s) is introduced into an appropriate host cell by any of a variety of suitable means, including transfection, electroporation or delivery by liposomes. DEAE-dextran, or calcium phosphate, may be useful in the transfection protocol.

After the introduction of the vector *in vitro*, recipient cells are grown in a selective medium, that is, medium that selects for the growth of vector-containing cells. Expression of the cloned gene sequence(s) results in the production of the heterologous protein.

According to the invention, this expression can take place in a continuous manner in the transformed cells, or in a controlled manner.

If desired, in *in vitro* culture, the expressed protein is isolated and purified in accordance with conventional conditions, such as extraction, precipitation, chromatography, affinity chromatography, electrophoresis, or the like.

The vectors obtained through the methods above, will provide sequences that, by definition, provide a transcriptional regulatory element of the invention (the serglycin promoter, and/or the enhancer element and/or the suppressor element). Such vectors may be designed with restriction enzyme sites that allow for the the insertion of a DNA sequence encoding a heterologous protein at a site or sites operably linked to the transcriptional regulatory complex (the promoter and any additional elements that alter promoter function).

Using the techniques described above, cotransfection of rat fibroblasts with  $\lambda$ MG-PG1 and the selectable marker pSV2 neo resulted in the establishment of fibroblast cell lines that had integrated both

foreign genes into their genome (Avraham *et al.*, *J. Biol. Chem.* 264:16719 (1989)). RNA blot analysis revealed that two of the rat fibroblast cell lines contained low, but detectable, levels of the 1.0-kb mRNA transcript that encodes mouse serglycin. No other gene that encodes a proteoglycan peptide core has been isolated and sequenced in its entirety. Neither has one been inserted into a foreign cell.

The ability of the transfected rat fibroblasts to transcribe the foreign mouse gene indicates that some of the regulatory elements in the gene's promoter are present in the isolated mouse genomic clone. When the 504-bp 5' flanking region of the mouse serglycin gene was compared to the corresponding 5' flanking region of the analogous human gene, a 119-bp region that immediately precedes the transcription-initiation site was found to be nearly identical. This nucleotide sequence is more highly conserved in evolution than any similar sized region of the gene that is translated into protein.

The 504-bp 5' flanking region of the mouse serglycin gene was linked to plasmid DNA that contains the structural sequences of the human growth hormone reporter gene, and the amount of growth hormone produced by different cell types transfected with the resulting plasmid construct quantified. With deletion analysis and site-directed mutagenesis, three motifs in the 5' flanking region of the mouse serglycin gene were identified that regulate its constitutive transcription. One of these elements suppressed transcription of the gene, whereas the other two elements enhanced its transcription. Due to the near identity of this 5' flanking region in the two species, it is likely the same *cis*-acting elements are used by all mouse and human cells that express this proteoglycan. As indicated by gel-mobility-shift assays, hematopoietic cells that transcribe the serglycin gene possess *trans*-acting factors in their nuclei that recognize these elements, and a different profile of *trans*-acting factors is present in fibroblasts that do not express the serglycin gene.

Using the enhancer that resides between nucleotide residues -118

-37-

and -81 as an enhancer, more growth hormone was produced in transiently-transfected cells than with control plasmid DNA containing a generic promoter. Because this *cis*-acting motif is one of the most potent enhancers now known for hematopoietic cells, it can be used as an effective tool to drive transcription (and thereby translation) of any foreign gene in hematopoietic cells.

#### IV. Characterization of the *Trans*-Acting Factors

Transcription is regulated by *trans*-acting factors that bind to distinct *cis*-acting elements usually located in the 5' flanking regions of genes, and these DNA-binding proteins can act in synergy to enhance transcription or in an opposing manner to suppress transcription. As assessed by gel mobility shift assays, rat basophilic leukemia-1 cells and rat-1 fibroblasts contain a number of DNA-binding proteins in their nuclei that specifically bind the region of DNA that contains the serglycin suppressor *cis*-acting element, the serglycin enhancer *cis*-acting element, and the proximal element of the serglycin promoter region. Based on their similar mobilities in the gel mobility shift assays, rat basophilic leukemia-1 cells and rat-1 fibroblasts contain a common *trans*-acting factor (B/F<sub>(-250/-161)</sub>-I) that binds to the suppressor element and a common *trans*-acting factor (B/F<sub>(-40/+24)</sub>-I) that binds to the proximal promoter. In addition, distinct *trans*-acting factors are present in each cell line. Rat-1 fibroblasts have distinct *trans*-acting factors that bind to the suppressor element (F<sub>(-250/-161)</sub>-II), the enhancer element (F<sub>(-118/-81)</sub>-I) and the proximal promoter (F<sub>(-40/+24)</sub>-II), whereas rat basophilic leukemia-1 cells have distinct *trans*-acting factors that bind to the enhancer element (B<sub>(-118/-81)</sub>-I) and the proximal promoter (B<sub>(-40/+24)</sub>-II).

Stably transfected rat-1 fibroblasts that have incorporated 10-20 copies of the mouse genomic clone  $\lambda$ -MG-PG1 into their genome constitutively express low levels of the 1.0 kb serglycin transcript. Based

on the transient transfections described herein, one reason normal fibroblasts contain no serglycin mRNA, and transfected fibroblasts contain only limited amounts, is due to the presence of the *trans*-acting factor of the invention in these mesenchymal cells, such *trans*-acting factor being  
5 very effective in suppressing transcription of this gene.

A computer search using the "Dynamics" program (Ghosh, D., *Nucleic Acids Res.* 18:1749-1756 (1990)) failed to reveal a conserved *cis*-acting element within residues -250 to -118 of the mouse and human serglycin gene that is recognized by a known suppressor DNA-binding  
10 protein, supporting the novelty of the *cis*-acting element present in this region of the serglycin gene. Because fibroblasts are more effective than rat basophilic leukemia-1 cells in their use of the element that resides between residues -250 and -190 to suppress transcription of the mouse serglycin gene, the responsible *trans*-acting factor may be more abundant,  
15 selectively expressed, or post-translationally modified to be more active. As assessed by the gel mobility shift assays with the residues -250 to -161 probe, the nuclear extracts of rat-1 fibroblasts contained at least one *trans*-acting factor that was not recognized in rat basophilic leukemia-1 cells.

#### V. Uses of the Invention

20 Although bacteria, yeast, and insect cells often can be transfected with foreign cDNAs or genes to obtain biologically-active recombinant proteins, in many situations it is necessary to express a protein in a mammalian cell so that it can be properly modified post-translationally. Some of the easiest cells to maintain in culture and to transfect with  
25 foreign DNA are immature hematopoietic cells such as rat basophilic leukemia-1 cells, mouse WEHI-3 monocytic cells, and mouse P815 mastocytoma cells. All three of these cell lines have their own spectrum of *trans*-acting DNA-binding proteins that bind to the *cis*-acting elements



of those genes that the cells are programmed to express. Thus, in order to obtain maximal expression of a foreign gene in a transfected cell, one must use the regulatory elements of a gene that is expressed in abundance in that cell type. Serglycin mRNA is expressed in abundance in a large number of immature mouse, rat, and human hematopoietic cells that are easy to maintain in culture. The cis-acting elements of the invention, in the 5' flanking region of the mouse serglycin gene can, for the first time, be used to drive transcription and translation of a foreign gene in transfected rat basophilic leukemia cells and WEHI-3 monocytic cells.

Thus, the invention is useful for the expression of any protein in a mammalian, and especially a hematopoietic, cell system, especially any protein that requires the mammalian environment for post-translational modifications, including glycosylation. Proteins of interest that may be so expressed include hormones, such as insulin and growth hormone, other peptide growth factors, cytokines, interferons, interleukins, enzymes, structural proteins, albumin, actin, etc., and especially c-kit ligand, granulocyte-macrophage colony stimulating factor, interferon- $\gamma$ , IL-1, IL-3, IL-4, IL-9, IL-10, nerve growth factor, and transforming growth factor- $\beta$ .

Many varieties of transcriptional control may be provided to a heterologous gene by the regulatory elements of the invention. In a host cell of hematopoietic cell origin, for example, if a genetic sequence encoding the 5' flanking region of the serglycin gene (for example, the proximal 504 bp of the 5' flanking region) is operably linked to a heterologous gene, such genetic sequence may be expected to express in a manner, and to a degree similar to that of the native serglycin peptide gene.

Expression of a desired heterologous protein in a host cell of hematopoietic cell origin may be achieved by operably linking genetic sequences encoding the enhancer element of the invention to a desired promoter sequence functional in such host cell, such element being located between nucleotides -118 and -81 of the serglycin gene, and such

-40-

element being dominantly active to stimulate transcription of operably linked genes in hematopoietic cells.

Expression of a desired heterologous protein in a host cell of hematopoietic cell origin may also be achieved by introducing genetic sequences encoding the unique and atypical eukaryotic promoter element operably linked to the coding sequence of the desired heterologous protein, such promoter element being located between nucleotides -40 and -20 of the serglycin gene, and such element being dominantly active for the promotion of transcription of operably linked genes in hematopoietic cells.

Expression in fibroblast hosts may be modified such that a desired gene that overexpresses an undesired protein in a fibroblast host may be "turned off" by introducing genetic sequences encoding the suppressor element of the invention, located between nucleotides -250 and -190 of the serglycin gene, on an integrating or viral vector that inserts such element into the transcriptional regulatory region of the gene.

For example, mast cell-derived glycosaminoglycans such as heparin and chondroitin sulfate di-B have potent biologic activities in different clinical situations. Unfortunately, prior to the invention, it has been difficult to obtain these glycosaminoglycans in sufficient quantity for analysis. Because of this problem, the biologic activities of mast cell-derived chondroitin sulfate E and chondroitin sulfate D have not even been tested. The ability to culture mast cells and, according to the invention, to alter the constituents of the mast cell's secretory granule using recombinant cytokines under the transcriptional control of the regulatory elements of the invention. Each cytokine or other desired factor may be examined for its ability to induce the polymerization of a specific type of glycosaminoglycan onto serglycin. Thus, mast cells may be induced to polymerize a specific type of glycosaminoglycan onto serglycin in response to specific recombinant cytokines, and, for the first time, the culture scaled up to obtain large amounts of the glycosaminoglycan of

-41-

interest. Alternatively, using recombinant technology an animal may be genetically altered such that it can be induced to express large numbers of that particular mast cell that contains the desired glycosaminoglycan.

Having now generally described the invention, the same will become better understood by reference to certain specific examples which are included herein for purposes of illustration only and are not intended to be limiting unless otherwise specified.

### EXAMPLES

#### Example 1

#### Construction and Screening of a HL-60 cDNA Library

The promyelocytic leukemia cell line, HL-60, is a transformed human cell that synthesizes chondroitin sulfate proteoglycans and stores these proteoglycans in its secretory granules. Under certain *in vitro* conditions, this cell can be induced to differentiate into cells that resemble neutrophils, monocytes, macrophages, eosinophils, and basophils. HL-60 cells (line CCL 240; American Tissue Type Collection, Rockville, Maryland, USA) were lysed in the presence of guanidine isothiocyanate (BRL, Gaithersburg, MD), and total RNA was purified by the CsCl density-gradient centrifugation technique of Chirgwin *et al.*, *Biochemistry* 18:5294 (1979). The poly (A)<sup>+</sup> RNA that was obtained by oligo (dT)-cellulose (Collaborative Research, Waltham, MA) chromatography (Aviv, H., and Leder, P., *Proc. Natl. Acad. Sci. USA* 69:1408 (1972)) was converted into cDNA (Okakajama, H., and Berg, P., *Mol. Cell Biol.* 2:161-170 (1982)). The resulting cDNAs were blunt ended with T4 DNA polymerase (Biolabs, Beverly, MA), the internal *EcoRI* sites methylated, and the cDNAs ligated to *EcoRI* poly-linkers. After selection of cDNAs of >500 bp by Sepharose CL-4B (Pharmacia) chromatography, the cDNAs were ligated to dephosphorylated  $\lambda$ gt10 DNA. *Escherichia coli*

-42-

(strain C600 Hfl) were infected with the resulting recombinant bacteriophages resulting in a library with a complexity  $>1 \times 10^6$ . The HL-60 cell-derived cDNA library was probed at 37°C with [ $\alpha$ - $^{32}$ P]dCTP (3000 Ci/mmol; New England Nuclear, Boston, MA) nick-translated pPG-1 in  
5 hybridization buffer (50% formamide, 5X SSC (0.15 M NaCl/15 mM sodium citrate), 2X Denhardt's buffer, 0.1% sodium dodecyl sulfate (SDS), 1 mM EDTA, 100  $\mu$ g/ml salmon sperm DNA carrier, and 10 mM sodium phosphate). The filters were washed at 37°C under conditions of low stringency of 1.0X SSC, 0.1% SDS, 1 mM EDTA, and 10 mM sodium  
10 phosphate, pH 7.0. Approximately 500,000 recombinants in the library were plated to isolate the clone designated cDNA-H4 (Figure 1). The HL-60 cell-derived cDNA library ( $\approx$ 500,000 recombinants) were re-screened using cDNA-H4 as the probe. Thirty clones that hybridized under conditions of high stringency (55°C; 0.2X SSC, 1 mM EDTA, 0.1%  
15 SDS, and 10 mM sodium phosphate, pH 7.0) with cDNA-H4 were isolated from the secondary screening of the library.

The individual HL-60 cell-derived cDNAs and their subcloned fragments were inserted into M13mp18 and M13mp19 (Amersham, Arlington Heights, IL) and sequenced by the dideoxy chain termination  
20 method of Sanger *et al.*, *Proc. Natl. Acad. Sci. USA* 74:5463 (1977). Both strands of cDNA-H4 were sequenced. The sequencing strategy is presented in Figure 1. The consensus nucleotide sequence of the HL-60 derived secretory granule proteoglycan peptide core cDNAs is shown in Figure 2.

25 A 249 bp *Eco*RI $\rightarrow$ *Eco*RI fragment has been isolated from a *Eco*RI digest of cDNA-H19. The nucleotide sequence of this fragment (Table I) contains the sequence expected for a polyadenylation site (underlined) and the poly(A) $^+$  tail (underlined). This fragment hybridizes to a genomic fragment that encodes the gene for this proteoglycan  
30 peptide core, and thus probably represents the next 249 bp of the transcript.

-43-

Table I

Consensus nucleotide sequence of the 3' end of the cDNA that encodes serglycin, the peptide core of the HL-60 cell secretory granule proteoglycan (Avraham, S. *et al.*, *Proc. Natl. Acad. Sci. USA* 86:3763-3767 (1989)).

GAATTCTTAA-AGGATTATGC-TTTAATGCTG-TTATCTATCT-  
TATTGTTCTT-GAAAATACCT-GCATT TTTTG-GTATCATGTT-  
CAACCAACAT-CATTATGAAA-TTAATTAGAT-TCCCATGGCC-  
ATAAAATGGC-TTTAAAGAAT-ATATATATAT-TTTTAAAGTA-  
GCTTGAGAAG-CAAATTGGCA-GGTAATATTT-CATACCTAAA-  
TTAAGACTCT-GACTTGGATT-GTGAATTATA-ATGATATGCC-  
CCTTTTCTTA-TAAAAACAAA-AAAAAAATAA-T [SEQ ID No. 1]

Example 2Chromosomal Localization of the Human Serglycin Gene

For the chromosome localization of the human gene that encodes cDNA-H4, DNA from five different human/mouse (lines 13C2, 24B2, 1711, 462TG, and 175) and 12 different human/hamster (lines 35A2, 35A4, 35B5, 35C1, 35D3, 35D5, 35E4, 35F1, 35F3, 35F5, 89E5, and 95A4) somatic cell hybrids were digested with *Bam*HI. The resulting fragments were resolved by agarose gel electrophoresis, and the DNA blots were analyzed under conditions of high stringency using cDNA-H4 as a probe. The percent discordance of the cDNA-H4 probe to each human chromosome was determined as described in Table II; a discordant fraction of 0.00 indicates that, in HL-60 cells, the serglycin gene is located on chromosome 10.

-44-

Table II

Segregation Pattern of cDNA-H4 with DNA from  
Human/Rodent Somatic Cell Hybrids

The DNA from different human/hamster and human/mouse somatic hybrid cell lines and the DNA from the controls were analyzed for their hybridization to cDNA-H4. The column designations are: +/+, both hybridization to cDNA/cDNA-H4 and the specific human chromosome are present; -/-, hybridization to the cDNA-H4 and the chromosome are both absent; +/-, hybridization is present but the chromosome is absent; and -/+, hybridization is absent but the chromosome is present. For calculation of the discordant fraction for each chromosome, the sum of the +/- and -/+ columns are divided by the sum of the +/+, -/-, +/-, and -/+ columns. The 19q+ category represents the der 19 translocation chromosomes for the hybrid clones derived from fusions with leukocytes from the two different X/19 translocation carriers. The X and Xq- categories represent the intact X and the der X translocation chromosomes. Bruns *et al.*, *Biochem Genet.* 17:1031-1059 (1979).

Human Chromosome	Segregation Pattern				Discordant Fraction
	+/+	-/-	+/-	-/+	
1	1	8	7	1	0.47
2	3	8	5	0	0.31
3	3	6	3	3	0.40
4	2	4	5	3	0.57
5	4	8	4	3	0.47
6	3	5	4	3	0.47
7	6	7	2	2	0.24
8	2	6	6	3	0.53
9	0	5	7	4	0.69
10	8	9	0	0	0.00
11	3	6	5	3	0.47
12	2	7	5	1	0.40
13	5	7	5	1	0.40
14	6	4	1	4	0.33
15	4	8	3	1	0.25
16	3	6	3	3	0.40
17	2	6	5	3	0.50
18	5	7	3	2	0.29
19 and 19q+	6	2	2	7	0.53

-45-

	20	3	3	3	6	0.60
	21	3	4	4	5	0.56
	22	3	6	2	3	0.36
	X and Xq-	2	4	6	4	0.63
5	Y	0	9	7	0	0.44

Example 3Identification of Nucleotide Sequences  
in the Human Genome that Encode Serglycin

The rat L2 cell-derived cDNA, pPG-1, disclosed in Bourdon *et al.*,  
 10 *Proc. Natl. Acad. Sci. USA* 82:1322 (1985) was used to identify the  
 genomic fragments encoding human serglycin from a *Bam*HI digest of  
 human genomic DNA. While no hybridization occurred when the DNA  
 blot was probed under conditions of high stringency with either pPG-1 or  
 pPG-M, or probed under conditions of low stringency with pPG-M, at  
 15 least 10 DNA fragments were visualized when the blot was probed under  
 conditions of low stringency with pPG-1. The large number of DNA  
 fragments detected suggested that there was a multi-gene family in the  
 human which contained repetitive sequences similar to those which  
 encode the serine-glycine repeat region of the L2 cell proteoglycan  
 20 peptide.

Example 4Isolation and Characterization of the Human Serglycin Gene

Subcloned fragments of the HL-60 cell-derived proteoglycan  
 cDNA, cDNA-H4, were radiolabeled with [ $\alpha$ -<sup>32</sup>P]dCTP (3000 Ci/mmol;  
 25 DuPont-New England Nuclear) to a specific activity of  $>10^8$  cpm/ $\mu$ g by  
 either nick translation (Maniatis, T. *et al.*, *Molecular Cloning: A  
 Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor,

New York, pp. 109-114 (1985)) or random priming (Feinberg, A.P. *et al.*, *Anal. Biochem* 166:224-229 (1983)), and then were used to screen  $\approx 10^6$  recombinants in a EMBL3 human genomic library (Klickstein, L.B. *et al.*, *J. Exp. Med.* 165:1095-1112 (1987)) by plaque hybridization.

5 Nitrocellulose filters (Millipore, Bedford, MA) were probed at 42°C in 50% formamide, 0.75 M NaCl, 75 mM sodium citrate, 5X Denhardt's buffer, 0.1% SDS, 1 mM EDTA, 100 µg/ml salmon sperm DNA carrier, and 10 mM sodium phosphate. The nitrocellulose filters were washed at 55°C with 30 mM NaCl, 3 mM sodium citrate, 0.1% SDS, 1 mM EDTA,

10 and 10 mM sodium phosphate, pH 7.0. Several independent clones were obtained using the entire 650 bp HL-60 cell-derived cDNA, cDNA-H4 (Figure 2). However, in order to obtain better representation of the 5' flanking region of the gene, the human genomic library was rescreened using the 136 bp 5'→*Kpn*I fragment of cDNA-H4 to isolate 2 additional

15 clones. The restriction maps of the clones were determined by incubating samples of their DNA separately with *Acc*I, *Bam*HI, *Eco*RI, *Hind*III, *Kpn*I, or *Sal*I (New England Biolabs, Beverly, MA). The digests were electrophoresed in 1% agarose gels, and the separated DNA fragments were transferred to Nytran membranes (Schleicher and Schuell, Keene,

20 NH) (Southern, E.M., *J. Mol. Biol.* 98:503-517 (1975)). The resulting DNA blots were probed with specific 5' (5'→*Kpn*I and 5'→*Xmn*I) and 3' (*Xmn*I→3' and *Acc*I→3') fragments of cDNA-H4.

#### Nucleotide Sequence Analysis of the Human Serglycin Gene.

Human genomic fragments (Figure 3) were subcloned into the Bluescript

25 (Stratagene, La Jolla, CA) plasmid vector using double enzyme polarized shotgun ligations to improve the efficiency of recombination and to maintain the orientation of the subclones (Kurtz, D.T. *et al.*, *Gene* 13:145-152 (1980)). Recombinant transformants were identified by colony hybridization and were restriction mapped by the same method as that

30 used above for the phage clones. Double stranded DNA sequencing



-47-

(Sanger, F. *et al.*, *Proc. Natl. Acad. Sci. USA* 74:5463-5467 (1977); Zhang, H. *et al.*, *Nuc. Acid. Res.* 16:1220-7 (1988)) was performed directly on the plasmid subclones using a Sequenase nucleotide sequencing kit (United States Biochemical, Cleveland, OH) and [ $\alpha$ -<sup>35</sup>S]dATP (1000 Ci/mmol; Amersham). Universal oligonucleotide primers (SK, KS, T3, T7 and M13rev; Stratagene) were used to determine the sequence of the first and last 300 nucleotides in each the subcloned fragment. Based on the nucleotide sequences of the sense strand and the antisense strand of the genomic fragment, two oligonucleotides that were each 18 nucleotides in length were synthesized on an Applied Biosystems 380A Oligonucleotide Synthesizer at the Harvard Microchemistry Facility, Cambridge, MA. These oligonucleotides were then used as primers to determine the contiguous nucleotide sequence of the next 200-250 nucleotides in each direction of the double stranded DNA. Additional oligonucleotides complementary to different regions of the cDNA were used as primers to extend the sequence from the exons in both directions. Nucleotide sequence data was entered and edited on an IBM-PC using the Clatech molecular biology software package. Database searches and homology comparisons with the mouse serglycin gene and other genes were performed using the computers at the Molecular Biology Computer Research Resource at Dana-Farber Cancer Institute, Boston, MA.

When a human genomic library was screened using the entire 650 bp cDNA-H4 probe, six independent clones were isolated (designated as  $\lambda$ HG-PG1 to  $\lambda$ HG-PG6). Restriction mapping of these clones revealed that all six of the clones lacked a 5' 12 kb *EcoRI*-->*EcoRI* fragment and failed to hybridize to a 136 bp 5'-->*KpnI* fragment of cDNA-H4. Rescreening of the genomic library with the 5' fragment of cDNA-H4 resulted in the isolation of two additional clones that were designated as  $\lambda$ HG-PG7 and  $\lambda$ HG-PG8, respectively. When analyzed by restriction mapping, clones  $\lambda$ HG-PG6,  $\lambda$ HG-PG7, and  $\lambda$ HG-PG8 contained

-48-

overlapping genomic sequences which taken together include the entire gene which encodes human serglycin.

Detailed restriction mapping of these subclones revealed that this gene spans at least 16.6 kb and consists of 3 exons (Figures 3 and 4).  
5 Between the first and second exon is a 8.4 kb intron, and between the second and third exon is a 6.7 kb intron. Both introns begin with the nucleotide sequence "GTAAG" and end with the sequence "CAG". Analysis of the nucleotide sequence of this gene revealed that exon 1 encodes the 5' untranslated region of the mRNA transcript and the entire  
10 27 amino acid hydrophobic signal peptide of the translated molecule. Exon 2 encodes a 49 amino acid portion of the peptide core (amino acid residues 28 to 76) which would be predicted to be the N-terminus of the molecule after the hydrophobic signal peptide is removed in the endoplasmic reticulum. Exon 3 (634 bp) is the largest exon and encodes  
15 the remaining 82 amino acids of the translated molecule and the entire 3' untranslated region of the mRNA transcript. These 82 amino acids encode a 17 amino acid sequence (residues 77 to 93) that immediately precedes the glycosaminoglycan attachment region, the 18 amino acid serine-glycine rich region (residues 94 to 111), the C-terminus of the  
20 translated molecule (residues 112 to 158).

Determination of the Transcription-Initiation Site of the Human Serglycin Gene. A S1 nuclease mapping analysis was performed to identify the transcription-initiation site of the human gene that encodes the peptide core of serglycin proteoglycans in HL-60 cells. A 4 kb *Sal*I -  
25 -> *Hind*III fragment of the genomic clone  $\lambda$ HG-PG7 was subcloned into Bluescript (designated pB5SH3). An oligonucleotide (5'-->CTTGAAGTGAAGGATTCCAGAA-->3' [SEQ ID No. 2]) was synthesized that corresponded to the residues 89 to 110 of the antisense strand of cDNA-H4. Ten nanograms of this oligonucleotide were hybridized to 4  $\mu$ g of  
30 alkali-denatured pB5SH3, and a complementary strand of DNA was

synthesized under conditions similar to that described above except that it was labeled with [ $\alpha$ - $^{32}\text{P}$ ]dATP. A 400 bp antisense DNA probe was isolated following electrophoresis of the synthetic product on a denaturing 8 M urea/8% polyacrylamide gel. The single-stranded radiolabeled DNA  
5 fragment was identified by autoradiography, electroeluted, and ethanol precipitated. A 50,000 cpm sample of the radiolabeled DNA fragment was hybridized to  $\approx 15 \mu\text{g}$  of HL-60 cell-derived total RNA (Chirgin, J.M. *et al.*, *Biochemistry* 18:5294-5299 (1979)) or  $1 \mu\text{g}$  of HL-60 cell-derived poly(A) $^{+}$  RNA (Aviv, H. *et al.*, *Proc. Natl. Acad. Sci. USA* 69:1408-1412  
10 (1972)) at  $48^{\circ}\text{C}$  for 16 h in 80% formamide, 400 mM NaCl, 1 mM EDTA, and 40 mM Pipes (pH 6.4). The  $^{32}\text{P}$ -DNA/RNA hybrid was incubated with 100 U of S1 nuclease (Pharmacia) for 60 min. At the end of the reaction, the sample was extracted with phenol, and ethanol precipitated at  $-80^{\circ}\text{C}$ . Three microliters of 1 mM EDTA and 10 mM Tris-HCl (pH  
15 8.0) and  $4 \mu\text{l}$  of formamide loading buffer were added to the precipitated sample. The sample was boiled, and loaded onto a 8 M urea/8% polyacrylamide sequencing gel along side a digest of  $^{32}\text{P}$ -labeled pBR322 (New England Biolabs) and a sequencing ladder of pBSH3 that had been primed with the same oligonucleotide. For two negative controls, S1  
20 nuclease reactions were concurrently preformed with  $15 \mu\text{g}$  of tRNA (Bethesda Research Labs) or MBBC (Razin, E. *et al.*, *J. Biol. Chem.* 257:7229-7236 (1982)) total RNA.

HL-60 cell-derived total RNA and poly(A) $^{+}$  RNA protected 132 nucleotides of the probe from degradation by S1 nuclease. Therefore, it  
25 was concluded that the putative transcription-initiation site in HL-60 cells for this gene resided 53 bp upstream of the translation-initiation site. The  $^{32}\text{P}$ -labeled 5' antisense 400 bp DNA fragment was not protected if it was incubated with tRNA or mouse mast cell RNA prior to exposure to S1 nuclease. This deduced transcription-initiation site in HL-60 cells  
30 corresponds to the deduced transcription-initiation site of the analogous gene that is expressed in BMMC-derived mast cells and rat basophilic

-50-

leukemia cells (Bourdon, M.A. *et al.*, *Mol. Cell Biol.* 7:33-40 (1987)), but not in rat L2 yolk sac tumor cells (Bourdon, M.A. *et al.*, *Mol. Cell Biol.* 7:33-40 (1987)).

5 Figure 3 is a restriction map of the human gene. Figure 4 is the nucleotide sequence of the gene that encodes human serglycin including 5' flanking and intron sequences.

Cloning of the Mouse Serglycin Gene. A 15 kb mouse genomic fragment containing the gene that encodes the mouse serglycin was cloned by screening a mouse genomic library derived from a *Sau3AI* digest of  
10 BALB/c mouse liver DNA (Avraham, S., *et al.*, *Proc. Natl. Acad. Sci. USA* 86:3763-3767 (1989)), using a [ $\alpha$ -<sup>32</sup>P]dCTP labeled 450 bp *AccI*->3' gene-specific fragment of a bone marrow-derived mast cells cDNA (cDNA-M6) that encodes the peptide core of mouse secretory granule proteoglycan using methods as described above. The nucleotide sequence  
15 and the deduced amino acid sequence of this gene is presented in Figure 5.

Neither the human nor the mouse gene have a classical TATA box (Breathnach, R. *et al.*, *Ann. Rev. Biochem.* 50:349-383 (1981)) or GC-rich element (Sehgal, A. *et al.*, *Mol. Cell Biol.* 8:3160-3167 (1988))  $\approx$ 30 bp  
20 upstream of its transcription-initiation site. Therefore, the serglycin gene that is expressed in hematopoietic cells has an unusual promoter. The 5' flanking region has not been described for any other human proteoglycan peptide core gene, and thus comparisons with genes that encode other proteoglycan peptide cores cannot yet be made. Of importance is the  
25 finding that 96% of the nucleotides that are present in a 119 bp nucleotide sequence just upstream of the transcription-initiation site of the human (residues -1 to -119) and mouse (residues -1 to -123) gene are identical. This degree of conservation greatly exceeds that obtained when any other 119 bp region within the exons of the gene in these two species

is compared, and suggests that important *cis* acting regulatory elements are present in this conserved nucleotide sequence.

### Example 5

#### Stable Transfection of Rat-1 Fibroblasts with the Mouse Serglycin Gene

Fisher rat-1 fibroblasts were grown in Dulbecco's modified essential medium (DMEM) supplemented with 10% fetal calf serum, 2 mM L-glutamine, 100 U/ml of penicillin, and 100  $\mu$ g/ml of streptomycin, at 37°C in a humidified atmosphere of 5% CO<sub>2</sub>. DNA cotransfections were performed essentially as described elsewhere (Southern, P.J. *et al.*, *J. Mol. Appl. Gen.* 1:327-341 (1982)). In brief, 3-4 x 10<sup>5</sup> rat-1 fibroblasts were placed into each 10-cm plastic culture dish containing DMEM for 12-24 h before cotransfection with the mouse genomic clone  $\lambda$ MG-PG1 and the selectable marker pSV2 neo. A calcium phosphate/DNA precipitate was created by adding 0.5 ml of a 250-mM solution of calcium phosphate containing 5  $\mu$ g of the  $\lambda$ MG-PG1 DNA and 0.5  $\mu$ g of pSV2 neo drop-wise in the presence of bubbling air to 0.5 ml of 280 mM NaCl, 10 mM KCl, 12 mM dextrose, 1.5 mM sodium phosphate, and 50 mM HEPES (pH 7.1). The precipitate that formed after a 30 min incubation was added to a culture dish of fibroblasts, and 10 to 18 h later, the DNA precipitate was removed. The transfected cells were washed twice with growth medium and then were allowed to recover for 24 h before being trypsinized and split at a ratio of 1:6. The resulting fibroblasts were plated into new 10-cm plastic dishes and cultured for 2 to 3 wk in DMEM containing 500  $\mu$ g/ml gentamicin (Gibco); the culture medium was changed every 3 days. At the end of this period, gentamicin-resistant colonies of transfected fibroblasts were individually picked with cloning cylinders and grown as cell lines in culture medium containing 100  $\mu$ g/ml gentamicin.

-52-

RNA and DNA Blot Analysis of Rat-1 Fibroblasts Stably Transfected with the Mouse Serglycin Gene. Total RNA was prepared from mouse bone marrow-derived mast cells, rat-1 fibroblasts, and transfected rat-1 fibroblasts by a guanidinium thiocyanate method (Chirgin, J.M. *et al.*, *Biochemistry* 18:5294-5299 (1979); Glisin, V. *et al.*, *Biochemistry* 13:2633-2637 (1974)). RNA (5 µg/lane) was electrophoresed in 1% formaldehyde-agarose gels, and transferred to Zetabind (Thomas, P.S., *Proc. Natl. Acad. Sci. USA* 77:5201-5205 (1980)). The resulting RNA blots were incubated at 42°C for 24 h in hybridization buffer containing a radiolabeled *AccI*→3' fragment of cDNA-M6. The blots were washed under conditions of high stringency, and autoradiography was performed. The mouse serglycin probe was removed from the blots by high temperature washing, and the blots were reprobed with an actin cDNA to quantitate the amount of mRNA that had been loaded in each lane.

DNA was isolated (Blin, N. *et al.*, *Nucleic Acids Res.* 3:2303-2308 (1976)) from the mouse liver, rat liver, rat-1 fibroblasts, and transfected rat-1 fibroblasts, and samples were digested (10 µg/digest) separately with *XmnI*, *BamHI*, *BglII*, *SspI*, *Sau3AI*, *HindIII*, or *EcoRI* for 4 h at 37°C. The fragments were resolved by agarose gel electrophoresis and were transferred to Zetabind. The resulting DNA blots were analyzed for hybridization under conditions of high stringency with the *AccI*→3' fragment of the mouse cDNA-M6 as a probe.

Expression of the Mouse Gene that Encodes the Peptide Core of Mouse Serglycin Gene in Transfected Rat-1 Fibroblasts. To demonstrate that λMG-PG1 contained the entire mouse serglycin gene, including its promoter region, and that this mouse genomic clone could be expressed in another mammalian cell, rat-1 fibroblasts were cotransfected with λMG-PG1 and the dominant neo-resistant selectable marker encoded by the plasmid pSV2 neo. Seventeen independent clones of neo-resistant transfected rat-1 fibroblasts were isolated and were expanded separately.

Total RNA was isolated from bone marrow-derived mast cells, neo-transfected rat-1 fibroblasts, and the cotransfected rat-1 fibroblast cell lines.

5 The gene-specific serglycin probe failed to hybridize to any transcript in RNA blots of non-transfected fibroblasts; however, it did hybridize to a 1.0-kb RNA transcript in mouse bone marrow-derived mast cells and in two of the cotransfected rat-1 fibroblast cell lines. Primer extension analyses were performed using RNA from the transfected fibroblasts to determine the transcription-initiation site. When RNA from 10 the transfected cells was used as an RNA template,  $\approx 80$  nucleotides were extended onto the oligonucleotide primer that corresponded to residues 78 to 98 of cDNA-M6, resulting in a DNA product of about 100 nucleotides in length. A DNA product of  $\approx 60$  nucleotides was obtained when the alternative primer that corresponded to residues 39 to 59 of 15 cDNA-M6 was used in the assay.

Genomic DNA was prepared from the above two clones of transfected rat-1 fibroblasts, and was digested with *Bgl*II, *Xmn*I, *Sal*I, or *Bam*HI. DNA blots of the digests were probed with the *Acc*I-->3' gene-specific fragment of cDNA-M6 to demonstrate that these transfected rat-1 20 fibroblast cell lines contained mouse serglycin genomic sequences. The mouse proteoglycan probe hybridized to a 2.7-kb fragment present in the *Bgl*II digest of mouse liver DNA, and to a 7.5-kb fragment in the *Bgl*II digests of both rat liver DNA and rat-1 fibroblast DNA. The transfected fibroblasts differed from the non-transfected rat-1 fibroblasts in that they 25 contained both the 2.7-kb and the 7.5-kb DNA fragments. Based on the relative intensity of hybridization of the gene-specific probe to the 2.7-kb fragment present in the *Bgl*II digests of equal amounts of mouse liver DNA and fibroblast DNA, the fibroblast cell lines may have incorporated 10-20 and 2-3 copies, respectively, of the mouse serglycin gene into their 30 genome.

-54-

Transfections have been performed in Chinese hamster ovary cells with a cDNA that encodes the peptide core of the fibroblast-derived dermatan sulfate proteoglycan called decorin (Yamaguchi, Y. *et al.*, *Nature* 336:244-246 (1988)) and in COS-7 cells with a cDNA that encodes the peptide core of the T-cell derived invariant chain proteoglycan that associates with Ia (Miller, J. *et al.*, *Proc. Natl. Acad. Sci. USA* 85:1359-1363 (1988)), but no transfections have been reported using a genomic clone that contains an entire proteoglycan peptide core gene.

#### Example 6

#### Preparation of Antibodies to Peptides of the Amino Acid Consensus Sequence Which Recognize Native HL-60 Cell Derived Glycin

A 16 amino acid peptide 02 [Ser-Asn-Lys-Ile-Pro-Arg-Leu-Arg-Thr-Asp-Leu-Phe-Pro-Lys-Thr-Arg] [SEQ ID No. 3] was chemically synthesized, coupled to hemocyanin, and injected into a New Zealand White rabbit. This peptide corresponds to residues 64-79 of the translated molecule and was a region of the core that preceded the serine-glycine rich glycosaminoglycan attachment region.

The induction of antibodies which specifically recognize the peptide core protein of human serglycin was tested as follows. The peptide (3 mg) was coupled with 5 mg of Keyhole Limpet hemocyanin (Sigma) in the presence of 0.25% glutaraldehyde, and polyclonal antibodies were raised to the coupled peptide in New Zealand White rabbits using standard immunization methodologies. Antibody titers in whole sera were measured using an enzyme linked immunosorbent assay (ELISA). Each microtiter well was incubated overnight at 4°C with 1 µg of synthetic peptide in phosphate buffered saline. After the remaining protein binding sites in the wells were blocked by a 1 h incubation with 1% (w/v) bovine serum albumin (Sigma), the wells were washed with phosphate buffered saline containing 1% (w/v) Tween 20. Rabbit sera that was serially



-55-

diluted in phosphate buffered saline was added, followed by horseradish peroxidase-conjugated goat anti-rabbit IgG (Bio-rad, Richmond, CA); the wells were then assayed spectrophotometrically for development of the 2,2'-azino-di-[3-ethyl-benzthiazoline sulfonate] dye (Boehringer-Mannheim, Indianapolis, IN). Peptide 01 (Ser-Val-Gln-Gly-Tyr-Pro-Thr-Gln-Arg-Ala-Arg-Tyr-Gln-Trp-Val-Arg) [SEQ ID No. 4] that corresponded to residues 24 to 39 of the deduced amino acid of cDNA-H4 was also synthesized and used in the ELISA to confirm the specificity of the rabbit antisera. Anti-peptide IgG was partially purified by ammonium sulfate precipitation followed by ion exchange chromatography.

Anti-peptide IgG (~30  $\mu$ g) was incubated with 100  $\mu$ l of a 15% (w/v) suspension of the Protein A-Sepharose beads (Sigma) in RIPA buffer for 1 h at room temperature. The resulting Protein A-Sepharose-IgG complex was added to 1 ml of RIPA cell lysates containing  $5 \times 10^6$  cell equivalents of [ $^{35}$ S]methionine-labeled or [ $^{35}$ S]sulfate-labeled HL-60 cells that had been precleared by incubation for 24 h with Protein A-Sepharose alone and then for 24 h with Protein A-Sepharose/preimmune IgG. After a 18-24 h incubation at 4°C with Protein A-Sepharose/anti-peptide IgG, the beads were washed 3 times by centrifugation with 0.1% bovine serum albumin, 0.5% Tween 20, and 10 mM phosphate buffered saline (pH 7.2) containing either 10 mM unlabeled methionine or unlabeled sodium sulfate. The bound radiolabeled antigens were eluted by suspending the beads in 60  $\mu$ l of Laemmli buffer and incubating for 5 min at 95°C. The eluates were electrophoresed in 15% SDS-PAGE gels, stained with Coomassie Brilliant blue, dried, and autoradiographed using Kodak XAR-5 film.

In the ELISA, the antiserum gave half-maximal binding at an approximate 500 fold dilution (Figure 6). The anti-peptide 02 serum failed to recognize peptide 01 which corresponded to deduced amino acid residues 24 to 39 of the same cDNA (Figure 6). The preimmune sera also failed to react with the coupled peptide 02. When 1  $\mu$ l of the

-56-

antisera was preincubated with 1  $\mu$ g of peptide 02 for 60 minutes at 25°C, no immunoreactivity was detected in the ELISA.

5 An IgG-enriched fraction of the anti-peptide 02 sera was used to determine if Protein A-Sepharose-bound antibodies would recognize the initially-translated serglycin and the mature proteoglycan. A prominent  
20,000  $M_r$  protein was specifically immunoprecipitated from lysates of 2 min [ $^{35}$ S]methionine-labeled HL-60 cells, whereas both a 20,000  $M_r$  protein and a macromolecule that barely entered the gel were specifically immunoprecipitated from 10 min radiolabeled cells. After a 10 min pulse  
10 and a 5 min chase, the 20,000  $M_r$  [ $^{35}$ S]methionine-labeled protein was less apparent while the macromolecule was somewhat increased. The [ $^{35}$ S]methionine-labeled macromolecule corresponded in size exactly to the [ $^{35}$ S]sulfate-labeled proteoglycan that was precipitated after an overnight radiolabeling of the cells with [ $^{35}$ S]sulfate. Because  
15 precipitation was inhibitable by preincubation of the Protein A-Sepharose-immune IgG with 1  $\mu$ g of the synthetic peptide 01, it was concluded that the rabbit anti-peptide 02 antibodies recognize the precursor and mature human glycin.

As shown in Figure 7, the size of the immunoprecipitated peptide  
20 core protein was approximately 13,000 daltons, consistent with the size predicted by Stevens *et al.*, *J. Biol. Chem.* 263:7287 (1988) for the peptide core which has lost its 27 amino acid signal peptide.

#### Example 7

##### Isolation of Serglycin Proteoglycans

25 Human serglycin proteoglycans can be isolated using common protein isolation techniques known in the art such as column chromatography, gel electrophoresis, affinity chromatography, or immuno-extraction techniques using the antibody described above. For example,

such proteoglycans may be extracted by the following procedure (Stevens, R.L., *et al.*, *J. Biol. Chem.* 260:14194-14200 (1985)).

Bone marrow-derived mast cells pellets are lysed by resuspension for 30 s in 50  $\mu$ l of 1% Zwittergent 3-12 containing protease inhibitors, followed by the addition of 2.35 ml of 4 M guanidine HCl (GnHCl) containing CsCl (density 1.4 g/ml). These detergent-GnHCl proteoglycan extracts are then pooled such that in a typical experiment 48 ml of extract is obtained from approximately  $2 \times 10^9$  bone marrow-derived mast cells, of which  $3 \times 10^7$  are radiolabeled. The pooled extracts are centrifuged at 17°C for 48 h at 95,000 X g, and the gradients are divided in most experiments into two equal fractions termed D<sub>1</sub> (bottom) and D<sub>2</sub> (top), respectively. The distribution of chondroitin sulfate E proteoglycan in fractions from the CsCl gradient or from subsequent ion exchange or gel filtration chromatography is determined by suspending a sample of each fraction in 12.5 ml of Hydrofluor and quantitating <sup>35</sup>S or <sup>3</sup>H radioactivity in the radiolabeled proteoglycan on a Tracor Analytic Mark III liquid scintillation counter. Protein is detected by the method of Lowry *et al.*, with bovine serum albumin as a standard or by optical density at 280 nm. Nucleic acids are detected at a wavelength of 260 nm. The bottom fraction of each CsCl gradient is placed in dialysis tubing of 50,000 M<sub>r</sub> cut-off and dialyzed at 4°C against 1 M NaCl for 24 h and then for an additional 24 h against 1 M urea containing 0.05 M Tris-HCl, pH 7.3. The dialysate is adjusted to 4 M in urea by the addition of solid urea and applied to a 0.8 X 29-cm column of DEAE-52 previously equilibrated in 4 M urea, 0.05 M Tris-HCl, pH 7.8. The ion exchange column is washed with 35 ml of 4 M urea, 0.05 M Tris-HCl, pH 7.8, and the chondroitin sulfate E proteoglycan eluted with a 180-ml linear gradient of NaCl (0-1.0 M) in the urea buffer at a flow rate of 4 ml/h. Two-ml fractions are collected, and the proteoglycan-enriched fractions, detected by monitoring a portion of the fraction for either <sup>35</sup>S or <sup>3</sup>H radioactivity if the cells have been prelabelled, are pooled, dialyzed 48 h at 4°C against 0.1 M

-58-

NH<sub>4</sub>HCO<sub>3</sub>, and lyophilized. This material is redissolved in 100  $\mu$ l of 4 M  
GnHCl/0.1 M sodium sulfate/0.1 M Tris-HCl, pH 7.0, applied to a 0.6 X  
100-cm column of Sepharose CL-4B in this same buffer, and eluted from  
the column at a flow rate of 1.5 ml/h. One-half ml fractions are collected  
5 and analyzed for radioactivity and absorbance at 280 nm. The proteo-  
glycan-containing fractions are pooled, dialyzed against 0.1 M NH<sub>4</sub>HCO<sub>3</sub>  
and lyophilized.

### Example 8

#### Isolation and Protease-Resistance 10 of HL-60 Cell Serglycin Proteoglycan

Radiolabeling of HL-60 Cells—HL-60 cells (line CCL 240;  
American Type Culture Collection, Bethesda, MD) were cultured in  
enriched medium [RPMI-1640 medium supplemented with 10% (v/v) fetal  
calf serum, 2 mM L-glutamine, 0.1 mM nonessential amino acids, 100  
15 U/ml of penicillin, and 100  $\mu$ g/ml of streptomycin (Gibco, Grand Island,  
NY)] at 37°C in a humidified atmosphere of 5% CO<sub>2</sub>. For [<sup>35</sup>S]  
methionine-labeling, HL-60 cells were preincubated at a concentration of  
10<sup>7</sup> cells/ml for 10 min in methionine-free, enriched medium containing  
dialyzed fetal calf serum. Approximately 500  $\mu$ Ci/ml of [<sup>35</sup>S]methionine  
20 (129 Ci/mmol; Amersham, Arlington Heights, IL) was then added. The  
HL-60 cells were incubated for an additional 2 to 10 min at 37°C,  
centrifuged in the cold at 120 x g, and washed at 4°C in enriched medium.  
In the pulse-chase experiments, HL-60 cells were [<sup>35</sup>S]methionine-labeled  
for 10 min, were washed as above, and were resuspended in normal  
25 enriched medium at 37°C for an additional 5 min. Aliquots of 5 x 10<sup>6</sup>  
[<sup>35</sup>S]methionine-labeled HL-60 cells were lysed in 1 ml of RIPA buffer  
[0.15 M NaCl, 1% deoxycholate, 1% Nonident P-40, 0.1% SDS, 10 mM  
N-ethylmaleimide, 2 mM phenylmethylsulfonyl fluoride, 10 mM NaF, and  
0.1 M Tris-HCl, pH 7.2], and immunoprecipitates of the lysates were

analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) as described below.

For [ $^{35}\text{S}$ ]sulfate labeling, HL-60 cells were incubated in enriched medium containing 50  $\mu\text{Ci/ml}$  of [ $^{35}\text{S}$ ]sulfate (-4000 Ci/mmol; DuPont-New England Nuclear, Boston, MA) for 1 h at a density of  $1 \times 10^7$  cells/ml or for 18 h at a density of  $2 \times 10^6$  cells/ml. The radiolabeled cells were centrifuged at 4°C for 10 min at 120 x g, and 150  $\mu\text{l}$  of 1% (w/v) zwittergent 3-12 (Calbiochem, San Diego, CA) containing 100  $\mu\text{g}$  of chondroitin sulfate A (Miles Scientific, Naperville, IL) and 100  $\mu\text{g}$  of heparin (Sigma, St. Louis, MO) glycosaminoglycan carriers were added to each cell pellet followed by 1.35 ml of 4 M GnHCl, 0.1 M sodium sulfate, and 0.1 M Tris-HCl. A sample of each lysate and supernatant was chromatographed on Sephadex G-25/PD-10 columns (Pharmacia, Piscataway, NJ) to quantitate the incorporation of [ $^{35}\text{S}$ ]sulfate into macromolecules.

In order to isolate the [ $^{35}\text{S}$ ]sulfate-labeled HL-60 cell serglycin proteoglycans, solid CsCl was added to the remainder of the cell lysates to achieve final densities of 1.4 g/ml. Following centrifugation for 48 h at  $\sim 100,000 \times g$ , the bottom 33% of each CsCl gradient was dialyzed sequentially against 0.5 M sodium acetate for 24 h and 0.1 M ammonium bicarbonate for an additional 24 h. The dialysates were lyophilized and redissolved in 0.4 ml of water. Samples of partially purified [ $^{35}\text{S}$ ]sulfate-labeled proteoglycans were incubated for 30 min with or without 10  $\mu\text{g}$  of Pronase (Calbiochem), and the digests were applied sequentially to a 0.8 x 85 cm column of Sepharose CL-6B (Pharmacia) that had been equilibrated with 4 M GnHCl, 0.1 M sodium sulfate, 0.1 M Tris-HCl, pH 7.2. As a control, samples of [ $^{35}\text{S}$ ]sulfate-labeled chondrosarcoma proteoglycans were analyzed in parallel for their susceptibility to Pronase by incubation of 1  $\mu\text{g}$  proteoglycan in 50  $\mu\text{l}$  Hanks' balanced salt solution at 37°C for 30 min with 5  $\mu\text{g}$  Pronase. Pronase-sensitive chondrosarcoma

-60-

proteoglycans are extracellular matrix proteins and are distinguished from secretory granule proteoglycans.

No substantial change in the hydrodynamic size of the HL-60 cell [<sup>35</sup>S]sulfate-labeled serglycin proteoglycans was detected following Pronase treatment, whereas rat [<sup>35</sup>S]sulfate-labeled chondrosarcoma proteoglycans were susceptible to degradation. These results show that the proteoglycan was resistant to Pronase digestion.

#### Example 9

#### Identification of Transcriptional Regulatory Elements of the Serglycin Gene and the Trans-acting Proteins That Bind To These Elements.

##### A. Experimental Procedures

1. Cell Lines — Rat basophilic leukemia-1 (RBL-1) cells (line CRL-1378, American Tissue Culture Collection (ATCC), Rockville, MD) and mouse myelomonocytic WEHI-3 cells (line TIB-68; ATCC) are cell lines of hematopoietic origin that abundantly express the serglycin transcript, whereas Fisher rat-1 fibroblasts (obtained from R.A. Weinberg, Whitehead Institute, Massachusetts Institute of Technology, Cambridge, MA) and mouse NIH/3T3 fibroblasts (line CRL 1658; ATCC) do not (Tantravahi *et al.*, *Proc. Natl. Acad. Sci. USA* 83:9207-9210 (1986)). Rat basophilic leukemia-1 cells, WEHI-3 cells, and the two fibroblast cell lines were grown in enriched medium (Dulbecco's modified Eagle's Medium supplemented with 10% fetal calf serum, 2 mM L-glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin (GIBCO, Grand Island, NY)) at 37°C in a humidified atmosphere of 6% CO<sub>2</sub>. Cells were split 1:4 every 3 days.

2. Plasmid DNA Constructs — With a polymerase chain reaction methodology (Saiki *et al.*, *Science* 239:487-491 (1988)), various lengths of the 504-bp 5' flanking region of the mouse serglycin gene

(Avraham *et al.*, *J. Biol. Chem.* 264:16719-16726 (1989)) that all extend upstream of residue +24 were obtained. DNA constructs were prepared by ligating the various DNA fragments into the *HindIII/XbaI* restriction-enzyme cloning sites of p $\phi$ GH and pSV40-hGH. p $\phi$ GH is a pUC12 plasmid that contains a promoterless hGH gene (Selden *et al.*, *Mol. Cell Biol.* 6:3173-3179 (1986)), and pSV40-hGH is a plasmid which contains the early SV40 promoter without its enhancer linked to the structural sequences of the hGH gene (Chung *et al.*, *Proc. Natl. Acad. Sci. USA* 83:7918-7922 (1986); Sarid *et al.*, *J. Biol. Chem.* 264:1022-1026 (1989)). A pUC12 plasmid that contains an enhancerless thymidine kinase promoter ligated to the hGH gene (pTKGH), and a pUC12 plasmid that contains both the enhancer and promoter of the mouse metallothionein-I gene ligated to the hGH gene (pXGH5) (Selden *et al.*, *Mol. Cell Biol.* 6:3173-3179 (1986)) were used as positive control plasmids in the DNA transfections. As described for other cell types (Sarid *et al.*, *J. Biol. Chem.* 264:1022-1026 (1989); Selden *et al.*, *Science* 236:714-718 (1987)), the latter well-characterized metallothionein-I-hGH fusion gene was used to optimize the DNA transfections and to normalize the efficiency of expression of hGH by the different cells.

Relevant 21-mer oligonucleotides that span the mutation site were used to perform site-directed mutagenesis (Zoller *et al.*, *DNA* 3:479-488 (1984)) on three nucleotides in a plasmid (designated pPG(-504/+24)hGH) containing the 504-bp 5' flanking region of the serglycin gene. In these constructs, the adenosine at residue -28 was converted to a cytosine, the cytosine at residue -30 was converted to an adenosine, or the adenosine at -38 was converted to a guanosine. The oligonucleotides used in the polymerase chain reactions, the site-directed mutagenesis, and the gel mobility shift experiments described below were synthesized on a Cyclone Plus DNA Synthesizer (Milligen/Bioscience, Novato, CA). The relevant nucleotide sequences within the different plasmid constructs was verified by dideoxy sequencing of the plasmid

-62-

DNA, as described by Sanger and coworkers (Sanger *et al.*, *Proc. Natl. Acad. Sci. USA* 74:5463-5467 (1977)) with modifications essential for double-stranded DNA sequencing (Chen *et al.*, *DNA* 4:165-170 (1985)).

3. Transfection Experiments — Rat basophilic leukemia-1  
5 cells were transiently transfected with the plasmid DNA constructs with DEAE-dextran (Lopata *et al.*, *Nucleic Acids Res.* 12:5707-5717 (1984)). The day before DNA transfection, the rat basophilic leukemia-1 cells ( $1 \times 10^6$ /dish) were plated in 100-mm culture dishes. Immediately before transfection, cells were washed once with 3 ml of serum-free enriched  
10 medium, and then 2.5 ml of serum-free enriched medium containing 5  $\mu$ g of supercoiled plasmid DNA complexed to 20 mg/ml DEAE-dextran ( $M_r$  500,000 daltons) was added. The dishes were incubated for 4 h at 37°C in a humidified atmosphere of 6%  $\text{CO}_2$ , the transfection solution was removed by aspiration, 3 ml of serum-free enriched medium containing  
15 10% dimethyl sulfoxide was added, and the cultures were incubated for 2 min more at room temperature. The transfected cells were treated with dimethyl sulfoxide, washed twice with 4 ml of serum-free enriched medium, and then cultured in 10 ml enriched medium at 37°C in a humidified atmosphere of 6%  $\text{CO}_2$ . No matter which DNA construct was  
20 used in the transfection, 100 later each culture dish contained approximately  $5 \times 10^6$  rat basophilic leukemia-1 cells.

Rat fibroblasts, mouse fibroblasts, and mouse WEHI-3 cells were transiently transfected with the plasmid DNA constructs using calcium phosphate (Avraham *et al.*, *J. Biol. Chem.* 264:16719-16726 (1989);  
25 Southern *et al.*, *J. Mol. Appl. Gen.* 1:327-341 (1982)). The cells ( $1 \times 10^6$ ) were suspended in 2.5 ml of enriched medium, 0.2 ml of a 250 mM solution of calcium phosphate containing 5  $\mu$ g of supercoiled plasmid DNA was added in a drop-wise manner, and the cultures were incubated overnight at 37°C in a humidified atmosphere of 6%  $\text{CO}_2$ . The  
30 transfection media were removed the following day, and the cells were washed and cultured at 37°C in 10 ml of fresh enriched medium in a



-63-

humidified atmosphere of 6% CO<sub>2</sub>. No matter which DNA construct was used in the transfection, 100 h later each culture dish contained approximately 4 x 10<sup>6</sup> fibroblasts. Approximately 100 h after the transfection by either method, 0.1 ml samples of culture medium were removed, and the levels of hGH were determined with an immunoassay kit from Nichols Institute Diagnostics (San Juan Capistrano, CA). The amounts of hGH in the culture media were determined by assaying the amounts of absorbed <sup>125</sup>I-labeled anti-hGH antibody in the sandwich assays.

The results of the transfection assays were normalized for transfection efficiency using pXGH5. The amount of hGH produced by pXGH5 transfected cells was arbitrarily assigned a value of 1, and then the relative amount hGH produced by each cell type transfected with the test plasmid was calculated as a ratio to that obtained with this control plasmid. In order to determine the promoter activity of various DNA constructs in dissimilar cell types transfected by different methods, a comparison of the relative amounts of hGH in the culture media for each cell type is preferable to a comparison of the absolute amounts of hGH (Sarid *et al.*, *J. Biol. Chem.* 264:1022-1026 (1989)). It has been reported in other studies (Selden *et al.*, *Mol. Cell Biol.* 6:3173-3179 (1986); Sarid *et al.*, *J. Biol. Chem.* 264:1022-1026 (1989); Selden *et al.*, *Science* 236:714-718 (1987)) that the amount of growth hormone produced by cells transfected with different hGH constructs is related to the amount of hGH mRNA in the transfected cells. To confirm that the variation in the amount of hGH in the culture media of cells transfected with the different constructs reflected a change in the level of hGH mRNA in the cells, total RNA was isolated from transfected rat basophilic leukemia-1 cells and rat-1 fibroblasts. Blots containing total RNA (10 µg/sample) were then prepared and probed with a <sup>32</sup>P-labeled 950-bp *Bgl*II/*Eco*RI fragment of the hGH cDNA present in pφGH.

4. DNA/Protein Binding Analyses — Nuclear extracts were prepared from rat basophilic leukemia-1 cells and rat-1 fibroblasts by a modification of the procedure described by Dignam and coworkers (Dignam *et al.*, *Nucleic Acids Res.* 11:1475-1489 (1983)). Each preparation of pelleted cells ( $10^8$ ) was washed once with 2.5 ml of ice-cold 10 mM Hepes (pH 7.9), 1.5 mM  $MgCl_2$ , 10 mM KCl, 1.0 mM dithiothreitol (DTT), 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 0.1% leupeptin, 0.1% pepstatin, and 0.1% aprotinin. After a 10-min incubation at 4°C in the same buffer, the cells were centrifuged for 3 min at 500 x g. The pelleted cells were resuspended in 1.0 ml of ice-cold buffer, lysed in a Dounce homogenizer, and centrifuged at 4°C for 10 min at 900 x g and then for 20 min at 16,000 x g. The supernatants were aspirated and the pelleted nuclear proteins were resuspended in 3 ml of 20 mM Hepes (pH 7.9), 25% glycerol, 1.5 mM  $MgCl_2$ , 0.42 M NaCl, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, 0.2% NP-40, 0.1% leupeptin, 0.1% pepstatin, and 0.1% aprotinin. The pellets were homogenized again in a Dounce homogenizer, agitated gently for 3 min, and centrifuged at 4°C for 30 min at 100,000 x g. The solubilized nuclear proteins in each supernatant were dialyzed at 4°C for 5 h against a 50-fold excess volume of 20 mM Hepes (pH 7.9), 20% glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, 0.1% leupeptin, 0.1% pepstatin, and 0.1% aprotinin, and then stored at -80°C in this buffer. The protein concentration of each nuclear extract was determined by the Bradford method (Bradford, M.M., *Anal. Biochem.* 72:248-254 (1976)) using a Bio-Rad (Richmond, CA) protein assay kit with bovine serum albumin as standard.

DNA/protein-binding experiments were carried out in 20  $\mu$ l of 10 mM Tris buffer (pH 7.5) containing 4% glycerol, 1 mM EDTA, 0.5 mM DTT, 0.1 M NaCl, 4  $\mu$ g of carrier poly(dI-dC) (Stratagene, La Jolla, CA), and 1 ng of a double-stranded  $^{32}$ P-end-labeled DNA probe corresponded to residues -250 to -161, residues -118 to -81, or residues -40 to +24 of the mouse serglycin gene (Avraham *et al.*, *J. Biol. Chem.* 264:16719-16726

(1989)). In the binding competition assays, 5 ng of the specific serglycin-derived unlabeled oligonucleotide, 100 ng of unlabeled sonicated salmon sperm DNA, or 100 ng of a 64-mer unlabeled double-stranded oligonucleotide that binds transcription factors NF1/CTF, SP1, AP1, and AP3 (Stratagene Catalog No. 203001) was added to each reaction. After incubation at 25°C for 30 min, gel mobility shift analyses were performed to detect the presence of specific DNA-binding proteins in the nuclear extracts. Samples were loaded onto a 5% non-denaturing polyacrylamide/bisacrylamide gel (30:1, w/w) that had been equilibrated before use by treatment for 1 h at 100 mA. The gels were run at 100 mA at 4°C until the bromophenol blue tracking dye ran approximately two-thirds the length of the gel. The gels were then dried under vacuum and autoradiographed generally for 16 to 24 h.

Two control mixing experiments were used to confirm that the 89-bp probe corresponding to residues -250 to -161 of the mouse serglycin gene bound to distinct *trans*-acting factors in nuclear extracts of rat basophilic leukemia-1 cells and rat-1 fibroblasts. In the first experiment,  $5 \times 10^7$  rat basophilic leukemia-1 cells and  $5 \times 10^7$  rat-1 fibroblasts were mixed together, and then a nuclear extract of the pooled cells was prepared and analyzed in the gel mobility shift assay. In the second experiment, nuclear extracts were prepared separately from rat basophilic leukemia-1 cells and rat-1 fibroblasts and were mixed in 1:1, 2:1, or 3:1 proportions just before being analyzed in the gel mobility shift assay.

#### B. Identification by Deletion Analysis of

-66-

Suppressors and Enhancers in the 5' Flanking  
Region of the Mouse Serglycin Gene

To determine if the proximal 5' flanking region of the mouse serglycin gene contains *cis* acting regulatory elements, a DNA fragment that extends 504 bp upstream and 24 bp downstream of the gene's transcription-initiation site was linked to p $\phi$ GH. Preliminary experiments were performed (analyzing the kinetics of secretion and the stability of hGH) to optimize the transfection assay in rat basophilic leukemia-1 cells, WEHI-3 cells, rat fibroblasts, and mouse fibroblasts. No hGH was detected in any cell pellet, and therefore apparently all translated hGH was secreted. Additional experiments revealed that hGH was not degraded following its secretion into the culture media.

Rat basophilic leukemia-1 cells, mouse WEHI-3 cells, rat-1 fibroblasts, and mouse 3T3 fibroblasts were transiently transfected with the resulting plasmid construct, designated pPG(-504/+24)hGH. The results of the transfection experiments were normalized for transfection efficiency relative to that obtained with the reference plasmid pXGH5. As additional controls, cells were transfected with pSV40-hGH and pTKGH. As shown in Table III, the relative amount of hGH present in the 4-d conditioned medium of transfected mouse WEHI-3 cells and rat basophilic leukemia-1 cells was 20- to 18-fold higher than for transfected fibroblasts of the respective species. Therefore, the 504-bp region immediately upstream of the transcription-initiation site of the mouse serglycin gene contains *cis*-acting elements that preferentially enhance transcription of this gene in hematopoietic cells.

-67-

TABLE III

Relative human growth hormone (hGH) production by four cell lines that have been transiently transfected with control plasmids and a plasmid that contains the 5' flanking region of the mouse serglycin gene fused to a promoterless human growth hormone gene.

Plasmid Construct	Relative Expression of hGH <sup>a</sup>				Ratio	
	Mouse Fib.	Mouse WEHI-3	Rat-1 Fib.	rat basophilic leukemia-1 cells	WEHI-3/ Mouse Fib.	rat basophilic leukemia-1/ Rat Fib.
pXGH5	1.0	1.0	1.0	1.0	1.0	1.0
pSV40-hGH	ND	ND	0.29 ± 0.06	0.47 ± 0.16	ND	1.6
pTKGH	0.30 ± 0.05	0.11 ± 0.04	ND	0.72 ± 0.20	0.3	ND
pPG(-504/+24)hGH	0.01 ± 0.02	0.20 ± 0.05	0.04 ± 0.01	0.72 ± 0.12	20	18

Fib., fibroblasts; ND, not determined; rat basophilic leukemia-1, rat basophilic leukemia-1 cells; and WEHI-3, mouse myelomonocytic cells.

<sup>a</sup> Results are expressed as the mean ± SD of 5 to 6 experiments of 4 d duration, with each experiment performed on 2 replicate dishes of cells.

To locate more precisely these *cis*-acting elements, 9 additional plasmid constructs were prepared that had progressive deletions of the 5' flanking region of this mouse gene fused to the hGH gene in p $\phi$ GH, as shown in Figure 8A. Rat basophilic leukemia-1 cells and rat-1 fibroblasts transfected with constructs pPG(-423/+24)hGH, pPG(-333/+24)hGH, and pPG(-250/+24)hGH produced amounts of hGH comparable to the corresponding cells transfected with pXGH5. The production of hGH was enhanced ~2.5-fold when rat basophilic leukemia-1 cells were transfected with construct pPG(-190/+24) hGH; production of hGH was also enhanced when rat basophilic leukemia-1 cells were transfected with the pPG(-118/+24)hGH construct. When rat-1 fibroblasts were transfected with constructs pPG(-190/+24)hGH and pPG(-118/+24)hGH, production of hGH increased 21-fold and 24-fold, respectively. Therefore, at least one *cis*-acting element resides between -250 and -190 that suppresses transcription of the serglycin gene in cells, and this negative element is dominantly active in fibroblasts. Although rat basophilic leukemia-1 cells and rat-1 fibroblasts transfected with constructs

pPG(-81/+24)hGH, pPG(-63/+24)hGH, and pPG(-40/+24)hGH produced some hGH, the amount was substantially less than that produced by cells transfected with construct pPG(-118/+24)hGH. Thus, at least one *cis*-acting element in the nucleotide sequence -118 to -81 constitutively enhances transcription of the serglycin gene in rat basophilic leukemia-1 cells and fibroblasts. When normalized for the efficiency of transfection, rat basophilic leukemia-1 cells transfected with construct pPG(-118/+24)hGH produced 2.7-fold ( $p < 0.05$ ) more hGH than similarly transfected fibroblasts, indicating that the enhancer between residues -118 and -81 is more dominantly active in rat basophilic leukemia-1 cells than in fibroblasts. Because no hGH was produced by rat basophilic leukemia-1 cells or rat-1 fibroblasts transfected with construct pPG(-20/+24)hGH, the proximal element in the promoter region of this gene must reside between -40 and -20.

As assessed by RNA blot analysis (Fig. 8B), rat basophilic leukemia-1 cells contained abundant amounts of hGH mRNA when transfected with pPG(-504/+24)hGH, pPG(-118/+24)hGH, pSV40-hGH, or pXGH5. A lesser amount of hGH mRNA was present in rat basophilic leukemia-1 cells transfected with pPG(-40/+24)hGH, and no hGH mRNA was detected in cells transfected with p $\phi$ GH. Rat-1 fibroblasts contained abundant amounts of hGH mRNA when transfected with pPG(-118/+24) hGH, pSV40-hGH, or pXGH5, but the amount was below detection when transfected with pPG(-504/+24)hGH or p $\phi$ GH (Fig. 8B). The level of hGH mRNA in rat-1 fibroblasts transfected with pPG(-40/+24)hGH was less than replicate fibroblasts transfected with pPG(-118/+24)hGH.

To determine if the positive *cis*-acting element in residues -118 to -44 of the mouse serglycin gene functions as an enhancer and to confirm that the negative *cis*-acting element resides upstream of the enhancer, two 5' flanking regions of the gene were ligated into pSV40-hGH to create constructs pPG(-250/-44)SV40-hGH and pPG(-118/-44)SV40-hGH (Fig.

-69-

9). Rat basophilic leukemia-1 cells and rat-1 fibroblasts were then transfected with pSV40-hGH or with one of the two plasmid constructs, and the relative levels of hGH in the culture media were determined. Greater than 3-fold more hGH was detected in the culture medium of rat  
5 basophilic leukemia-1 cells and rat-1 fibroblasts transfected with pPG(-118/-44)SV40-hGH relative to pSV40-hGH, indicating the enhancer in this 5' flanking region of the gene functions as an enhancer. This regulatory element also induced rat basophilic leukemia-1 cells and rat-1 fibroblast to produce ~3.6-fold more hGH when linked in the plasmid in  
10 its opposite orientation 2.6 kb upstream of the SV40 early promoter (Fig. 9). The finding that the level of hGH produced by fibroblasts transfected with pPG(-250/-44)SV40-hGH is approximately one-half that of fibroblasts transfected with pPG(-118/-44)SV40-hGH again indicates that there is a negative *cis*-acting element within the more distal 5' flanking region of the  
15 mouse serglycin gene that is active in fibroblasts.

C. Identification of the Proximal Region of the Promoter of the Mouse Serglycin Gene by Site-Directed Mutagenesis — Although no classical TATA box is present ~30 bp 5' of the transcription-initiation site of the mouse or human serglycin gene, a ACCTCTTTCTAAAGGG  
20 [SEQ ID No. 5] sequence is present beginning 22 nucleotides upstream of the transcription-initiation site. Site-directed mutagenesis was performed to determine whether or not this region was part of the proximal promoter of the gene. Constructs were prepared in which the adenosine at -28 of pPG(-504/+24)hGH was converted into a cytosine, the  
25 cytosine at residue -30 was converted into an adenosine, or the adenosine at -38 was converted into a guanosine (Table IV). Relative to cells transfected with pPG(-504/+24)hGH, substantially less hGH was produced by rat-1 fibroblasts and rat basophilic leukemia-1 cells transfected with any one of the three mutated constructs. The greatest  
30 inhibition occurred with the construct that had a mutated residue -30.

-70-

TABLE IV

Relative human growth hormone (hGH) production by rat basophilic leukemia-1 cells and rat-1 fibroblasts transfected with constructs containing a normal and mutated proximal promoter region of the mouse serglycin gene.

Nucleotide	Mutation	Relative Expression of hGH <sup>*</sup>	
		rat basophilic leukemia-1 cells	Rat-1 fibroblasts
Sequence	Position		
ACCTCTTTCTAAAAGGG (native) [SEQ ID No. 5]	None	1.00	1.00
ACCTCTTTCTCAAAGGG (mutated) [SEQ ID No. 6]	-28 bp	0.32 ± 0.04	0.38 ± 0.06
ACCTCTTTATAAAAGGG (mutated) [SEQ ID No. 7]	-30 bp	0.09 ± 0.02	0.17 ± 0.03
GCCTCTTTCTAAAAGGG (mutated) [SEQ ID No. 8]	-38 bp	0.43 ± 0.02	0.69 ± 0.02

\* Results are expressed as the mean ± 1/2 range of two experiments with each experiment performed on 8 replicate dishes of cells.

Protein/DNA binding Analyses - Gel mobility shift assays were used to determine whether or not rat basophilic leukemia-1 cells and rat-1 fibroblasts contain *trans*-acting factors in their nuclei that bind specifically to the three identified *cis*-acting regulatory elements in the 5' flanking region of the mouse serglycin gene. An 89-bp <sup>32</sup>P-labeled DNA fragment containing the putative suppressor and corresponding to residues -250 to -161 of the mouse serglycin gene was gel electrophoresed before and after it had been incubated with the nuclear extracts from rat basophilic leukemia-1 cells and rat-1 fibroblasts. As shown in Figure 10 for one of four experiments, in the absence of nuclear extract, the radioactive probe migrated to its expected position at the bottom of the gel (lane 1). When the <sup>32</sup>P-labeled probe was incubated with the nuclear extracts from either one of the two populations of cells before electrophoresis, it was selectively retained in the gel by a putative DNA-binding protein (designated B/F<sub>(-250/-161)</sub>-I) (lanes 2 and 5). The binding of this <sup>32</sup>P-labeled oligonucleotide to B/F<sub>(-250/-161)</sub>-I was specific because retention of the probe was diminished when a 5-fold excess of the same nonradioactive DNA fragment was included in the assay (lanes 3 and 6):



retention was not diminished when a 100-fold excess of sonicated salmon sperm DNA (lanes 4 and 7) or the 64-mer oligonucleotide that binds NF1/CTF, SP1, AP1, and AP3 was included in the assay. Based on its differential mobility in this gel mobility shift assay, a second *trans*-acting factor ( $F_{(-250/-161)}\text{-II}$ ) was detected only in the nuclear extracts of rat-1 fibroblasts. Because rat basophilic leukemia-1 cells, but not fibroblasts, contain proteases in their secretory granules (one of which is a chymase active at neutral pH (Seldin *et al.*, *Proc. Natl. Acad. Sci. USA* 82:3871-3875 (1985))), two control experiments were carried out to determine if the absence of  $F_{(-250/-161)}\text{-II}$  in rat basophilic leukemia-1 cells was a consequence of the isolation procedure used to obtain the nuclear DNA-binding proteins. When the nuclear extracts from rat basophilic leukemia-1 cells and fibroblasts were mixed in different proportions before analysis, the level of  $F_{(-250/-161)}\text{-II}$  detected in the gel mobility shift assay varied according to the amount of fibroblast-derived nuclear protein that was used in the assay. Furthermore, the amounts of  $B/F_{(-250/-161)}\text{-I}$  and  $F_{(-250/-161)}\text{-II}$  detected in the nuclear extract of a pooled preparation of rat basophilic leukemia-1 cells and fibroblasts were compatible with the results of the extracts of the individual cell types. Thus, the absence of  $F_{(-250/-161)}\text{-II}$  in the nuclear extracts of rat basophilic leukemia-1 cells was not a consequence of preferential proteolysis of this *trans*-acting factor in rat basophilic leukemia-1 cells.

A 37-bp  $^{32}\text{P}$ -labeled DNA fragment containing the putative enhancer element and corresponding to residues -118 to -81 of the mouse serglycin gene was gel electrophoresed before and after it had been incubated with the nuclear extracts from rat basophilic leukemia-1 cells and rat-1 fibroblasts. As shown in Figure 11 for one of 3 experiments, a retarded species ( $B_{(-118/-81)}\text{-I}$ ) was obtained when the  $^{32}\text{P}$ -labeled oligonucleotide was incubated with rat basophilic leukemia-1 cell-derived nuclear extracts (lane 2).  $B_{(-118/-81)}\text{-I}$  possessed a mobility different from that of the retarded species,  $F_{(-118/-81)}\text{-I}$ , present in the nuclear extracts of

-72-

rat-1 fibroblasts (lane 5). The ability to inhibit binding of the  $^{32}\text{P}$ -labeled oligonucleotide to  $\text{B}_{(-118/-81)}\text{-I}$  and to  $\text{F}_{(-118/-81)}\text{-I}$  with a 5-fold excess of the same nonradioactive DNA probe (lanes 3 and 6), but not with a 100-fold excess of sonicated salmon sperm DNA (lanes 4 and 7) or the 64-mer oligonucleotide that binds NF1/CTF, SP1, AP1, and AP3 (data not shown), indicated that these interactions were specific. Although an additional retarded species was observed in this gel mobility shift assay when either nuclear extract was used, its binding could not be inhibited by an excess of the specific nonradioactive oligonucleotide.

10 A 64-bp  $^{32}\text{P}$ -labeled DNA fragment containing the putative proximal promoter element and corresponding to residues -40 to +24 of the mouse serglycin gene was also gel electrophoresed before and after it had been incubated with the nuclear extracts from rat basophilic leukemia-1 cells and rat-1 fibroblasts. As shown in Figure 12 for one of 15 3 experiments, when the  $^{32}\text{P}$ -labeled probe was incubated with the nuclear extracts from either one of the two populations of cells before electrophoresis, a new species, designated  $\text{B}/\text{F}_{(-40/+24)}\text{-I}$ , was observed that migrated more slowly in the gel (lanes 2 and 7). The binding of this  $^{32}\text{P}$ -labeled oligonucleotide to  $\text{B}/\text{F}_{(-40/+24)}\text{-I}$  could be competitively 20 inhibited by a 5-fold excess of the same nonradioactive DNA probe (lanes 3 and 8), but not by a 100-fold excess of sonicated salmon sperm DNA or the oligonucleotide that binds NF1/CTF, SP1, AP1, and AP3. Additional distinct *trans*-acting factors, designated  $\text{F}_{(-40/+24)}\text{-II}$  and  $\text{B}_{(-40/+24)}\text{-II}$ , were detected in the nuclear extracts of rat-1 fibroblasts and rat basophilic 25 leukemia-1 cells, respectively. The binding of this 64-bp serglycin proteoglycan -derived  $^{32}\text{P}$ -labeled probe to  $\text{F}_{(-40/+24)}\text{-II}$ ,  $\text{B}/\text{F}_{(-40/+24)}\text{-I}$ , and  $\text{B}_{(-40/+24)}\text{-II}$  was minimally diminished in the competition assay when nonradioactive DNA that had a mutated residue -28, -30, or -38 was used in a 50-fold excess over the nonmutated  $^{32}\text{P}$ -labeled probe (Fig. 12).

Example 10Methylation of the Human Serglycin Gene

The serglycin gene-derived *Alu* sequences were aligned with the *Alu* consensus nucleotide sequence of Jurka, J. *et al.*, *Proc. Natl. Acad. Sci. USA* 85:4775-4778 (1988); their locations and characteristics are depicted in Figure 13A and Table V.

Table V  
Distribution and Type of *Alu* Elements in the Human Serglycin Gene

Twenty-one *Alu* elements were detected in the nucleotide sequence of the human serglycin gene (Figure 13A). Of these, 10 were identified in the introns. Thirteen were of the S type, and 8 were of the J type. In two instances, only approximately one-half of an *Alu* element was inserted in the gene. These elements were oriented in the sense (F) or anti-sense (R) direction relative to the rest of the human serglycin gene.

	Number	Location	Type	Direction
	1	5'-flanking region	S	R
	2	5'-flanking region	S	F
	3	Intron 1	S	F
20	4	Intron 1	S	F
	5	Intron 1	S	F
	6	Intron 1	S	F
	7	Intron 1	J	R
	8	Intron 1	S	F
25	9	Intron 1	S	F
	10	Intron 1	J	F
	11	Intron 2	J	R
	12	Intron 2	S	R
	13	Intron 2	J(1/2)	F
30	14	Intron 2	J	F
	15	Intron 2	J	R
	16	Intron 2	J(1/2)	F
	17	Intron 2	S	F

-74-

18	Intron 2	S	R
19	Intron 2	J	F
20	Intron 2	S	R
21	Intron 2	S	F

---

5

Fifteen diagnostic positions were examined to determine if the *Alu* elements were of the "S" or the "J" subfamily. Because the J subfamily of *Alu* elements is more similar to 7SL RNA than the S subfamily, the J type probably is a more primitive *Alu* element. Eleven of the *Alu* elements were of the S type, whereas eight were of the J type. The two *Alu* elements in the 5' flanking region were of the S type. The *Alu* elements were present in both orientations. Thirteen were oriented in the sense direction of the gene's exons; whereas the other six were orientated in the anti-sense direction.

15

The deduced nucleotide sequence of the human serglycin gene was used to determine the methylation pattern of this gene in cells that do and do not transcribe it. Because of their hybridization to corresponding regions within other genes, it was not possible to probe genomic DNA blots with short DNA fragments of the human serglycin gene that contained *Alu* DNA sequences. Thus, knowledge of the exact location of the *Alu* repetitive elements within the serglycin gene (Figure 13A) permitted the avoidance of those sequences in the methylation study. Because HL-60 cells, but not Molt-4 cells, contain serglycin mRNA, DNA was isolated from these two cell types and the methylation patterns of their serglycin genes determined. The location of all of the sites susceptible to *HpaII* and *MspI* within the human serglycin gene were determined (Figure 13B) and PCT methodology was used to construct 13 probes (designated A-M) to determine how many of these 5'-CCGG-3' sequences contained an internal 5' methylcytosine.

20

25

-75-

Human genomic DNA was prepared as described by Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., pp. 9.16-9.19 (1989), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. HL-60 cells and Molt-4 cells were each centrifuged at 1500 x g for 10 min at 4°C. The supernatants were removed, and the cells were washed twice with ice-cold 0.14M NaCl, 2.7 mM KCl, 25 mM Tris, pH 7.4. The cells were suspended in 10 mM Tris, pH 8.0, containing 0.1 M EDTA, 0.5% SDS, and 20 µg/ml pancreatic RNase and were incubated for 1 hour at 37°C. After Proteinase K (100 µg/ml) treatment under standard conditions, the digests were extracted 4 times with Tris-saturated phenol and then with chloroform. The genomic DNAs in the resulting solutions were precipitated with ethanol. Samples containing about 10 µg of DNA either were dissolved in 10 mM MgCl<sub>2</sub> and 20 mM Tris, pH 7.4, and digested with *Hpa* II (GIBCO-BRL), or were dissolved in 10 mM MgCl<sub>2</sub> and 50 mM Tris, pH 8.0, and digested with *Msp*I (GIBCO-BRL). These two restriction enzymes both cleave the unmethylated nucleotide sequence 5'-CCGG-3', but only *Msp*I cleaves this sequence if the internal C is 5-methylcytosine (Waalwijk, C., and Flavell, R.A., *Nucleic Acids Res.* 5:3231-3236 (1976)). The digests were electrophoresed in 1% agarose gels and transferred (Southern, E.M., *J. Mol. Biol.* 98:503-517 (1975)) to Duralon membranes (Stratagene). The resulting DNA blots were hybridized with random-primed, PCR-derived, 250-880-bp probes that correspond to 13 different regions of the human serglycin gene. These DNA probes were generated with either the HL-60 cell-derived serglycin cDNA (Stevens *et al.*, *J. Biol. Chem.* 263:7287-7291 (1988)) or the serglycin genomic subclones (Nicodemus *et al.*, *J. Biol. Chem.* 265:5889-5896 (1990)) as templates. The probes used in this methylation study were specific to the human serglycin gene because each hybridized to a single genomic fragment.

When blots containing digested genomic DNA from HL-60 cells were analyzed with the intron 1 probes A-F, each probe hybridized to a

-76-

DNA fragment in the *MspI* digest that was identical in size with the corresponding fragment in the *HpaII* digest. Therefore, the 5'-flanking region and intron 1 of the serglycin gene were both hypomethylated in HL-80 cells. Like the intron 1 probes, the intron 2 probe J hybridized to identically sized fragments in the *HpaII* and *MspI* digests of genomic DNA from HL-60 cells. In contrast, several of the other sites in intron 2 of the serglycin gene were at least partially methylated in HL-60 cells. Which of the five *HpaII/MspI* sites at the 3' end of this gene are methylated could not be conclusively determined because of their proximity to one another, but probes K and M both hybridized to larger DNA fragments in the *HpaII* digest, as compared with the *MspI* digest. Thus, some, if not all, of these sites at the 3' end of the serglycin gene are methylated in HL-60 cells. Whereas probes G, H, and I all hybridized to the expected size of DNA fragments after digestion of genomic DNA with *MspI*, they hybridized to two to three fragments after digestion with *HpaII*. The presence of a 3.2-kb DNA fragment that hybridizes to both probe H and probe I argues that the *HpaII* sites that reside at 11.5 and 11.8 kb in the serglycin gene are methylated in most HL-60 cells. In a second experiment, these two sites were methylated in almost all of the HL-60 cells in the culture. Exon 2 probe G hybridized to two approximately equal fragments in the *HpaII*-digest, indicating that the *HpaII/MspI* site at 9.5 kb in the serglycin gene was methylated in approximately 50% of the HL-60 in the culture. These findings are not the result of incomplete digestion of the DNA samples of nonspecific hybridization of the probes with a fragment from another gene because the same blot yielded single bands after hybridization with other probes and because single bands were detected when *MspI*-digested genomic DNAs were analyzed with these same serglycin-derived probes.

In contrast to the gene in HL-60 cells, the serglycin gene in Molt-4 cells was highly methylated. Probing with any of the PCR-derived DNA fragments yielded DNA genomic fragments of >10 kb, indicating that

-77-

most, if not all, of the *HpaII* sites in the serglycin gene of Molt-4 cells were methylated.

Several genes have been reported preferentially to contain *cis*-acting regulatory elements in their first introns. Although it has not been determined if the transcription-regulatory activities of any of these elements in intron-1 are effected by methylation, it has been shown that CpG methylation of the cAMP response element found in the promoters of many genes abolishes its transcriptional regulatory activity (Iguchi-Ariga, S.M.M. *et al.*, *Genes & Dev.* 3:612-619 (1989)). The diminished methylation of the first intron of the serglycin gene in a cell that contains abundant levels of this transcript, but not in a cell that does not transcribe the gene, suggests that specific methylation-dependent nucleotide sequences in intron 1 act in concert with the identified sequences in the 5'-flanking region (Avraham, S. *et al.*, *J. Biol. Chem.* 267:610-617 (1992)) to regulate transcription of the serglycin gene in different cell types.

All references cited herein are incorporated herein fully by reference. Having now fully described the invention, it will be understood by those with skill in the art that the scope may be performed within a wide and equivalent range of conditions, parameters and the like, without affecting the spirit or scope of the invention or any embodiment thereof.

-78-

## WHAT IS CLAIMED IS:

1. An isolated DNA sequence consisting essentially of the 5' regulatory region of the human or mouse serglycin gene, said regulatory region comprising:
  - 5           a. the promoter element located between nucleotides -40 and -20 of said serglycin gene;
  - b. the negative transcriptional regulatory element located between nucleotides -250 and -190 of said serglycin gene; and
  - 10          c. the positive transcriptional regulatory element located between nucleotides -118 and -81 of said serglycin gene.
2. The sequence of claim 1, wherein said serglycin gene is the human gene.
- 15          3. The sequence of claim 2, wherein said sequence consists essentially of bases -250 through -1 of Figure 4A.
4. The sequence of claim 1, wherein said serglycin gene is the mouse gene.
5. The sequence of claim 4, wherein said sequence consists  
20 essentially of bases -250 through -1 of Figure 5.
6. An isolated DNA sequence consisting essentially of the negative transcriptional regulatory element located between nucleotides -250 and -190 of the serglycin gene, and such element being dominantly active to inhibit transcription of operably linked genes in fibroblast hosts.



-79-

7. The sequence of claim 6, wherein such element is from the mouse serglycin gene.

8. The sequence of claim 7, wherein such element consists essentially of bases -250 through -190 of Figure 5.

5 9. The sequence of claim 6, wherein such element is from the human serglycin gene.

10. The sequence of claim 9, wherein such element consists essentially of bases -250 through -190 of Figure 4A.

10 11. An isolated DNA sequence consisting essentially of the enhancer transcriptional regulatory element located between nucleotides -118 and -81 of the serglycin gene, and such element being dominantly active to stimulate transcription of operably linked genes in hematopoietic cells.

15 12. The sequence of claim 11, wherein such element is from the mouse hematopoietic serglycin gene.

13. The sequence of claim 12, wherein such element consists essentially of bases -118 through -81 of Figure 5.

14. The sequence of claim 11, wherein such element is from the human serglycin gene.

20 15. The sequence of claim 14, wherein such element consists essentially of bases -118 through -81 of Figure 4A.

-80-

16. An isolated DNA sequence consisting essentially of the eukaryotic promoter element located between nucleotides -40 and -20 of the serglycin gene, and such element being dominantly active for the promotion of transcription inoperably linked genes in hematopoietic cells.

5           17. The sequence of claim 16, wherein such element is from the mouse serglycin gene.

18. The sequence of claim 17, wherein such element consists essentially of bases -40 through -20 of Figure 5.

10           19. The sequence of claim 16, wherein such element is from the human serglycin gene.

20. The sequence of claim 19, wherein such element consists essentially of bases -40 through -20 of Figure 4A.

21. A DNA expression vector comprising the DNA sequence of any of claims 1-20.

15           22. The expression vector of claim 21, wherein said vector is a plasmid.

23. The expression vector of claim 22, wherein said vector is an *E. coli*/mammalian cell shuttle vector.

20           24. The expression vector of any of claim 21, wherein said DNA sequence is operably linked to a gene of interest.

25. A host cell transformed with the DNA expression vector of claim 21.

-81-

26. A host cell transformed with the DNA expression vector of claim 24.

27. The host cell of claim 24, wherein said host is a hematopoietic cell.

5 28. The host cell of claim 24, wherein said host is a fibroblast.

29. A method for producing a gene of interest, said method comprising:

- 10 (a) transforming a host cell with the expression vector of claim 24, wherein said gene of interest is operably linked to said DNA sequence and heterologous to said DNA sequence;
- (b) expressing said gene of interest in said host cell; and
- (c) collecting said gene of interest.

30. A method for inhibiting the production of a gene, said method comprising:

- 15 (a) transforming a host cell with the expression vector of claim 24, wherein said gene of interest is operably linked to said DNA sequence and wherein said gene of interest encodes an antisense RNA complementary to said gene whose production it is desired to inhibit; and
- 20 (b) expressing said antisense RNA in said host cell.

31. A cell-free composition comprising B/F<sub>(-250/-161)</sub>-I.

32. A cell-free composition comprising B/F<sub>(-250/-161)</sub>-II.

-82-

33. A cell-free composition comprising  $B_{(-118/-81)}-I$ .
34. A cell-free composition comprising  $F_{(-118/-81)}-I$ .
35. A cell-free composition comprising  $B/F_{(-40/+24)}-I$ .
36. A cell-free composition comprising  $F_{(-40/+24)}-II$ .
- 5 37. A cell-free composition comprising  $B_{(-40/+24)}-II$ .

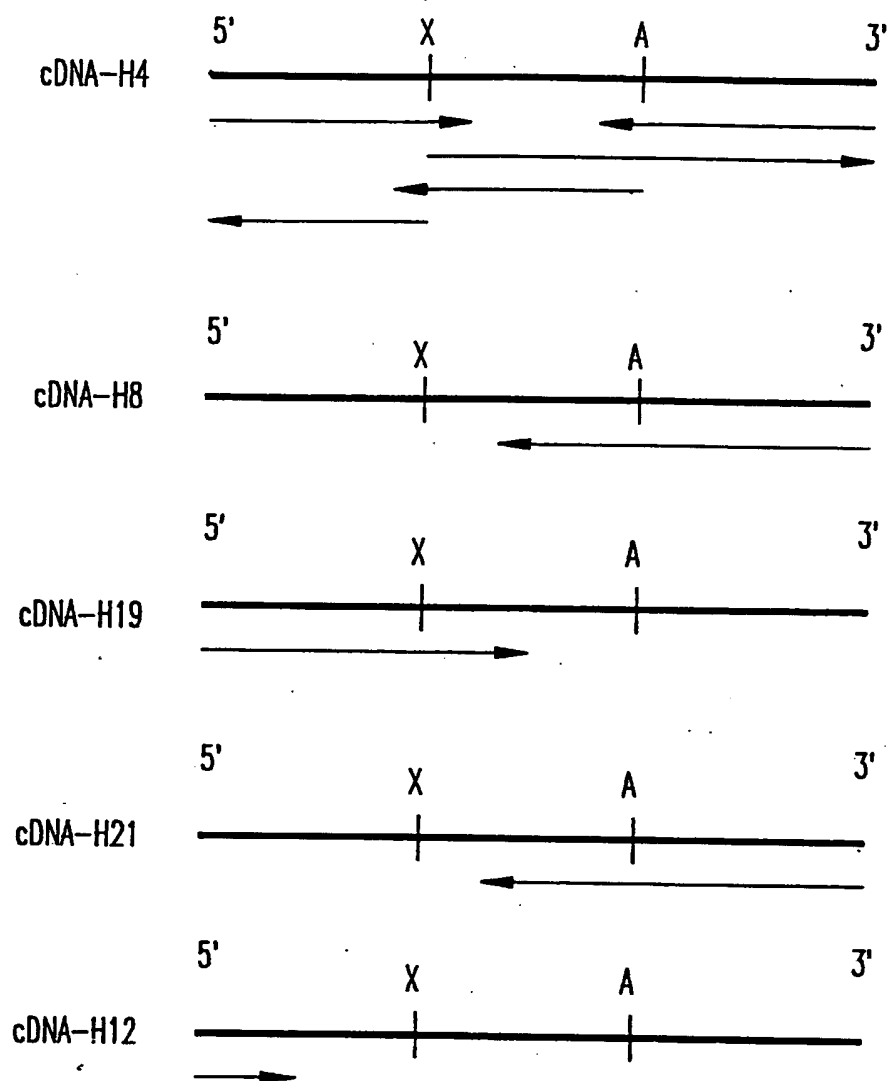
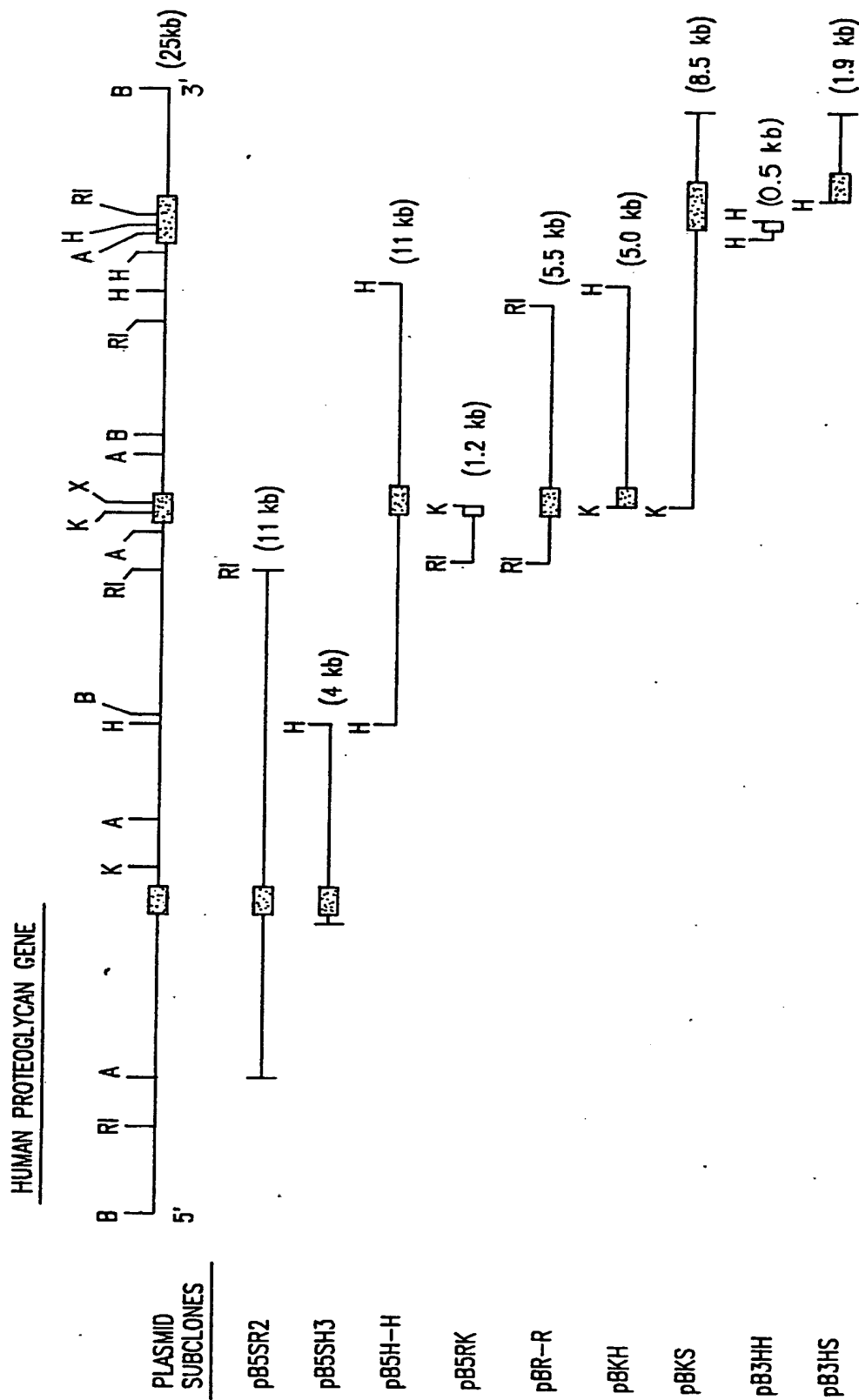


FIG.1

**SUBSTITUTE SHEET**



**FIG. 3**

4/25

-621 GCCACTGCTCTCCAGCCTGGGTGACAGAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAA  
 -557 AAAAAAGAGAAAAAGAGAGAGAGAACTGTTTCATCTGAAATCCGACAACTCATTCTTGAAAGTTAGAGCTCAGC  
 -479 TTTGAAGTTTCACCTTCACGAGCTTGGCTCAGTGAGGTATGTTACTCCCCGGTAAAAAGAAAAATGAAGAGAATGTTTT  
 -401 ATGTTGAAAGTGCTTGGTGACGAAAGGCAGCACCTAGATCCCTTATCTCATAAAAAATGCAGCAGATTCTTAATATT  
 -323 AGCAATCTAGTATTTAGATTGTTACCTGAAGAAAGGAAAAACAACTGTCCCAATGCTGATTCTACTGTTTCGGTGG  
 -246 GAAAAAAAAATGTCTTGCAAGCAAGTGGCAAAACAACAACTTTTGAAGAGCAGGCCCTGGGGGGAGTCCAGTACAGT  
 -168 TTCATAATGGGTATGAATAGTTATTTTACTGTGTTCCCCCACCCTTTCTTTCTGGGTTTTGATGTGGATGTCCTT  
 -89 CTATTTGTTCAAGAAATTGTGACGTGTGTTCTGGGCAGGGTTTGAGGTTTGGAACTTTCTAAAAGGGACAGAGAG

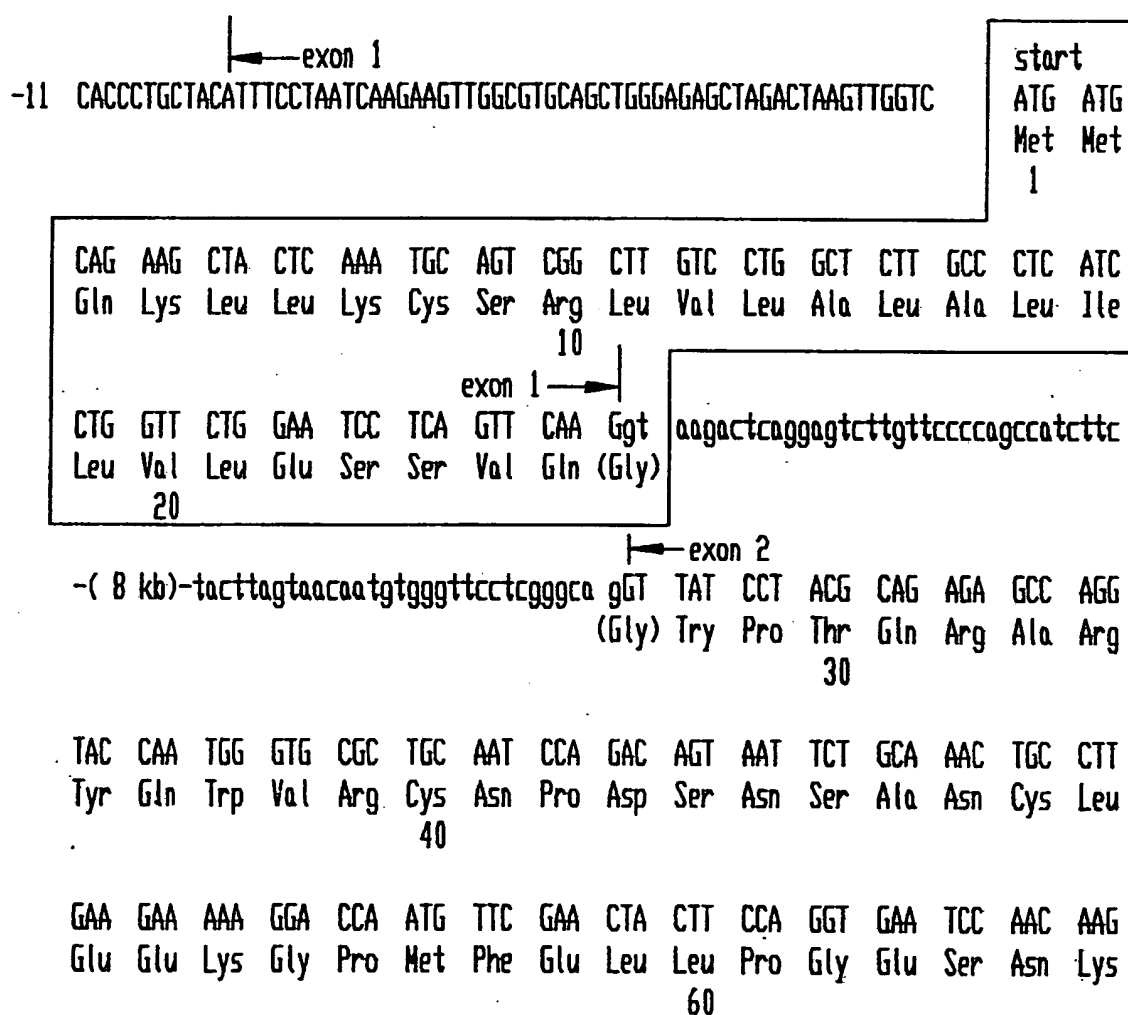


FIG.4A



5/25

exon 2 →

ATC CCC CGT CTG AGG ACT GAC CTT TTT CCg taagtggacttttctctaattaattaatt  
 Ile Pro Arg Leu Arg Thr Asp Leu Phe (Pro)  
 70

← exon 3

-(6 kb)-tccactggttttttccatttttctttcatacttc agA AAG ACG AGA ATC CAG GAC  
 (Pro) Lys Thr Arg Ile Gln Asp  
 80

TTG	AAT	CGT	ATC	TTC	CCA	CTT	TCT	GAG	GAC	TAC	TCT	GGA	TCA	GGC	TTC
Leu	Asn	Arg	Ile	Phe	Pro	Leu	Ser	Glu	Asp	Try	Ser	Gly	Ser	Gly	Phe
90															

GGC	TCC	GGC	TCC	GGC	TCT	GGA	TCA	GGA	TCT	GGG	AGT	GGC	TTC	CTA	ACG
Gly	Ser	Gly	Ser	Gly	Ser	Gly	Ser	Gly	Ser	Gly	Ser	Gly	Phe	Leu	Thr
100											110				

GAA	ATG	GAA	CAG	GAT	TAC	CAA	CTA	GTA	GAC	GAA	AGT	GAT	GCT	TTC	CAT
Glu	Met	Glu	Gln	Asp	Try	Gln	Leu	Val	Asp	Glu	Ser	Asp	Ala	Phe	His
120											130				

GAC	AAC	CTT	AGG	TCT	CTT	GAC	AGG	AAT	CTG	CCC	TCA	GAC	AGC	CAG	GAC
Asp	Asn	Leu	Arg	Ser	Leu	Asp	Arg	Asn	Leu	Pro	Ser	Asp	Ser	Gln	Asp
140															

TTG	GGT	CAA	CAT	GGA	TTA	GAA	GAG	GAT	TTT	ATG	TTA	TAA	AAGAGGATTTTC		
Leu	Gly	Gln	His	Gly	Leu	Glu	Glu	Asp	Phe	Met	Leu	stop			
150												160			

CCACCTTGACACCAGGCAATGTAGTTAGCATATTTATGTACCATGGTTATATGATTAATCTTGGGACAAAGAATTTT  
 ATAGAAATTTTAAACATCTGAAAAAGAGCTTAAGTTTATCATCCTTTTTTCTCATGAATTCCTAAAGGATTAT  
 GCTTTAATGCTGTATCTATCTTATTGTTCTTGAAAAACCTGCATTTTTTGGTATCATGTTCAACCAACATCATTAT  
 GAAATTAATTAGATTCCCATGGCCATAAAATGGCTTTAAAGAATATATATATATTTTTAAAGTAGCTTGAGAAGCAAA  
 TTGGCAGGTAATATTCATACCTAAATTAAGACTCTGACTTGGATTGTGAATTATAATGATATGCCCTTTTCTTATA  
 AAAACAAAAAAATAAT

FIG.4A(cont.)

SUBSTITUTE SHEET

GCCACTGCTCTCCAGCCTGGGTGACAGAGTGAGACTC  
-584 CATCTCAAAAAAAAAAAAAAAAAAGAGAAAAAGAGAGAAAGAACTG  
TTCATCTGAAATCCGACAACTCATTCTTGAAGGTTAGAGCTCAGCTTTGAAGTTCACTT  
CAGGAGCTTGGCTCAGTGAGGTATGTTACTCCCGGTGAAAAAGAAAATGAAGAGAAATGT  
-404 TTTATGTTGAAAGTGCTTGGTGACGAAAAGGCAGCACCTAGATCCCTTATCTCATAAAAA  
ATGCAGCAGATTCTTAATATTAGCAATGTAGTATTTAGATTGTTACCTGAAGAAAGGAAA  
AACAACTGTCCCAATGCTGATTCTACTGTTTCGGTGGGAAAAAAAAATGCTTGCAGG  
- 224 CAAGTGGCAACAACAAACTTTTGAAGAGCAGGCCTGGGGGAGTCCAGTACAGTTTC  
ATAATGGGTATGAATAGTTATTTTACTGTGTTCCTCCACCCCTTTCTTTCTGGGTTT  
GATGTGGATGCTTTCTATTTGTTGAGGAAATTGTGACGTGTGTTCTGGGCAGGGTTGA  
-44 GGTTTTGAACATTTTCTAAAAGGGACAGAGAGCACCTGCTAC

1  
ATTTCCTAATCAAGAA  
GTTGGCGTGAGCTGGGAGAGCTAGACTAAGTTGGTCATGATGCAGAAGCTACTCAAATG  
CAGTCGGCTTGTCTGGCTCTTGCCCTCATCCTGGTCTGGAATCCTCAGTTCAAG

EXON 1

GTAA  
137 GACTCAGGAGTCTTGTTCCTCCAGCCATCTTCTCTGTAAGCCCTGTGGTCCATGCAAGTCA  
TTATATTCATTTTAAGGCATAGAAATGTATAATTTGTGAGAAAGGAGGCAAGAGAAAGG  
ATTTGGGGTCCGTGAACCTTTAATATGAGTTCTGTTAAGTTTGGTACCAAGAAAAATTA  
317 AACTCTGTGGCGTGTGCAGTCTTGAAACTCTTACAATGATTGAAATGTGCTATTTTGGG  
ATGAAAATGTGAGTTTATAAATTTAAAGCTCAAAAAGGAATCTAGAAAATGACTCC  
TGTGCCCTGTTGCATGGAGGAGATGGCACCTTTGACTGTTGGGGGGTGTCTGCCACCCCT  
497 AAGTGTCTACATCAGCCCCAAGTTTGTAGTGGCTGTGACGGTGTGATTGTTATTTAACA  
CTGGGAGACGTTATATTCCAATTGGGGTGAATCTGACTGTGTGATTTTCTTTCTTTT  
TTTTTTTTTAAAGATAAACTTGGTTCTTACTGAAAACCAATTATGGTTAGACATAGTTC  
677 ATGTAACCTCTCAGATTTTAAAGAGAGGCAATAATTTGGTATTTGTGCTCTTGCT  
CAGAGAAGCATCATATTCGGAATATCTTCTAGGTTATCTACCATTTAGTGTGTTTA  
GTCAGACTGAAACAACTTAAACCTGTAATGACTAAGACAATGAAAATGATAGGCTTGTA  
857 AGAAAAATACAATTTGTTATTTCTTGGCAATAAGGAATCATGTCTAAATAGACGGAGG  
TCATGGCTTGATAGAGAGATGGCTGAACCTATAGTAGAAAAACACTAGGTTCCGCCAAT  
GGTAAGGGAAATGTTGAGTCACAATGACACACATGTCCTAGATTTGTTTCGTCAAAGCGA

FIG.4B

SUBSTITUTE SHEET

1,037 CTTTGGTTGTCATGATCTTACTTCCGGTGGAGATGAAATCTTACAGATGATCGCAGAGA  
CATTCATTTTATGTTGGAAATTTATAAAATCATTTTCTTCTAGTTATGCTAATGCTGAAA  
AAAGAGCAAGTAATGTTTCTGGAACGTTATTAATTTATGATTTTTAAATATAAAACAT  
1,217 TGTCAATTGTAGGGAACAGGCTTCACTGGGATCTTTAGGGAATATCTTCAGCTTGATGA  
AATAATCCCGAATAGCCAAGTGGTCTGACAAGATCGAGAGTAATGAGGCCATACTTTA  
GTACAGTCTTGAATGGCCAGATGGTCTGGGCATACCCCAACCAGAGATATGTAAGTCTT  
1,397 TATGTTGTCAAAATTTCCAGAAACATGAATTTCCCACTAAGATTCATTAAGGAAAATA  
GAATGAAAACAAAACGTTCTTGATAATATTCATTAGAAAAGAAATGAAGAAGGCCGGG  
CATGGTGGCTCAGCCTGTAAATCCAGCACTTTGAGAGGCCAAGGTAGGCAGATCATGAG  
1,577 GTCAGGAGTTTGAGACCAGCCTGGCCAACATAGTGAATCCCGTCTCTACCAAAAATACA  
AAAAATTAGCCGGGCATGGTGGCACACCTGTCATCCAGCTACTCAGGAGGCTGAGG  
CAGGAGAATTGCTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCTGAGATTGCACCACTGT  
1,757 ACTACAGCCTAGGTGACAGTGCAAGACTCTGTGAGAAAGAAAGAGAGAGAGAGAGA  
AAGAAAGGAAAGAAAGAAAGAGAGGAAAGAAATAATTATCATGAAATTGTATAGAAT  
ACTAGCATTATGTCATGACCTCGTAGGTTAGCTCTTTGTTAGAAAAGGAAACCATAGA  
1,937 AAGAGACAAGGGAGAACTGACAAACTAGGGTGTTCGAAAAAAGGCTCTCAGTATCGG  
GCTCAAGGGCTTGTGCCACATCTGAGCATGCAGGGAATAGATGTCCCCACTGGCTGC  
ACATGTGAGTACTGCCGCACAAGGCTGTGATGTGAAGAGTCATGACACCATTTCTCAC  
2,117 ACCTCCACGCAATGCCAGATATGATTCCACAACATTCTTCTGTCTTATAAAAGTGTTA  
TCTAGCCCGTTGGTTTGGCAGATGAAATCAACTAGGCTTTTGGCTTGCTTTTACTGAGCA  
TATTCAAAACCATTTCAAGTCACTATAGTGGTTTGTCTGGGTTGCCATAACAAAGTACCA  
2,297 CGGACTGAGTGGCTTAAATAACAGAAATGTATTTCTGACAGTTCGAGATGGGAGTCC  
AAGATCAAGGTGCTGGCAGGGTCGATCGCATTCGAGGCCCTCTCTCTTGGCTTATAGAT  
GGCGCCTTCTCCCTTGTCTGCACATGGCCTTTCCTCCATGCATCCGTGTCTTAATACCAT  
2,477 CTTCTTAGAGGGTCACCAGGCATTGGATCAGGGCCACCCTAATGGCCTCATCTTAACACTAC  
CTATATCTGCAATGACCTATTTCTGAACAATTCACATTGTGAGATTCTGTGGGTTAGA  
ACTGGAACATATGAATTTGGTGGTGGTATATTTTATTATAAGTCAAACCAAGTAAAGA  
2,657 TGTGGGGTAAGATTGTGTTTACCAAGCACAAAGAAATGGAAATTTGGGGATGTGTAACCTC  
TGGAGAGCACAGATGACTAATCTATTTAATGTAGGGCTCCAGGGGATTTGATGAGGCTG  
TGAATCTCCACTTTTATGCTCTCTTTTCCAATGACACCCATAAAGAAAAAAATGGA  
2,837 ATATCCATGAACAGGTGCAGCCAAGGAGGCCAGGCCGCCATGTGTCCACTGTATACTGT  
CTCCTAGCTCACAGGAATGATACTGATCCACTCCTTGTGCTGCTCTTTGTAAGTGATTT  
CACATCCATTCTCTGGTAATCATCATCACATTCCCTGTGATGAGGAATTAGCACCATT

FIG.4C

SUBSTITUTE SHEET

3,017 TAGAGGAGAAAACGGATCTGACATTTCTCATCTCATTGCTCTATACATTAACCTCTTG  
CAAAAAATTTGTGAGTCTTGCCCAAGACCCATTACAACATAATTAACGGCTGAACCTGGTCG  
TCTGATTTCAAGGCCAGAAATTAACCTTTCTACTGCAGCTCATGGATCAGAGGTTTTCTTTA  
3,197 TTTAAACAAACAAACAAAAATCCTTTGACCGTAGCCCTTGCTATAACATTTCCCACTGA  
GTTGAGGGAGAAATTGAAAGTAACTTAGGAGCTTTTTATAGCTTGCTCAAACCATGCAAG  
AGTGGGGGAGCTTATCCATCTTGTGGAATTGATAGACCAGGAGGAACTAACTCCGGCTT  
3,377 AGATAATGCTACCATTTTGAATAAATCAAATGGTCTTCTTTTCCCTTCATGGTAGTTGC  
TGCTTAAGTTTCTTAACATGCCCTGCAGTAAGTTCCATTAAGAATAGGAAATTAGGCTC  
GGTACAGTGGCTCAGCTCTGTAGTCTTAGCACTTTGGGAGACCGAGGAGTGTGGATCACT  
3,557 TGAGGTCAGGAGTTCAAGACCAGCCTGGCCACATGGTGAAACCCCATCTCTACTAAAA  
ATACAAAAATTAGCCAGGCATAGTGGCATGCACCTGTAAATCCAGCTACTCGAGAGGCTG  
AGGCAGGAGAACTCACTTGAACCAAGGAGATGGAGGTTGCAGTGAGCCAAGATCATGCCAC  
3,737 TGCACTCCAGCCTGGCCGATAGACTGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAGA  
AAAAAGAAAAAGAAAAAAAAAGAAATTAGTATATTGTGATTATGTTGAGGGAAAAAGTTAG  
TACCATAATATAAAAAGGTATGGACTATTGGAGAAAGTTGTTTGCTTTGGTAACATTTAC  
3,917 TCATAGAAAGTATTTTGGTAAAGCAGGACTCAGGGTGGTGGGGAGGTGGGCAGTGAGGG  
ATAGGATTCAAATAAAAACCATTTTCCCTTGGAAATCCACTACACAATTAACCAACAAA  
TCCCATAAAGTGGACCTTTTAGGAAGATAACATTTCTATCCATGAGCATAGCCACTATAAT  
4,097 CACAAGACATTTATCTCAAGCAAGATAGAGTCAAGATACTCTCACAACCTCAGGGGCTGG  
AACTGTAAATTTTACATCCTGCCAACACCTTGAATAGCTATGTCAAGAAATTTAGTGTG  
TGTAACCTGTTCTTTATTTTAAAGTACATTTAACATCATCGGCCCAAATTAGATAGGCT  
4,277 TTTGGAGTGGGATCCCTTCTACTTTTGATTTCTTTATAAAATTTTAAATAGCTTTGTG  
AGATAGTGTTCACATACCATACAGTTCACCCATTAAGTGTTCATTCAGTGATTAGGC  
CAGGTGTGGTGGCTCATGCCCTGTAAATCCAGCATTTTGGGAGGCCAAGGCAAGTGGATCA  
4,457 CTTGAGGTCAAGAGTTTGAGACTAGCCTGGCCAACATGGTGAGAACTTGTCTCTACTAAA  
AATACAAAAATTAACCTGGGCATGGTGGTGTGCACCTGTAAATCCAGCTACTTGGGAGGCT  
GAGGCAGGAGAACTCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCTGAGATTGCACCA  
4,637 CTGCACTCCAGCCTGGATGACGGAGCAAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAG  
TGTTCAATGTTTTTATGATATTCACAGAGTCATGCAACCATCACTATAATTGCTTCTAG  
AACATTTTCATCATCCCCAAAGAAAGCCTTCGTTACGAATTTTAATTAGCTGAGATTCT  
4,817 TGAACCTCTGGGGAAATTTGTATTCTAGAAATATTTTACTAATATGCTACAGTTGTAT  
TTGTCATGCTGGTGAAAAGATGTGGTCTTTCACCTGGATGCTTTCTCATTAAAGCATTATT  
TTTCTGTTTAGCTTCCTGTGTGAGCAACATTTTCTCAGCTTGATACTCAGTGATCAGC

FIG.4D

SUBSTITUTE SHEET

9/25

4,997 GGCTTCAGAGAGACTGCCTAGGCCTGCTCTGTCCAGTACGGTAGCCACAAGTCACTTG  
TGGCTACTGAACACTTGAAATGTGGCCAAGGCAAATTGGGACAGGCTGTGAATGCAAAAT  
ATACAAGATTGGAAACCTTAGTATGAAGAAAAGAATGCAAAATATCCAGTAATAACTT  
5,177 TTACATTGATGATATGTTAAAGGACAATATTTGAATATGTTAGGTTAAATAAAATTAATT  
TCACCTGTTTCTTTTACTTTTAAAAATATGGCCGCTGGAACATTTAAACTCCCTATGT  
GGTTTGCTTTGTGTTTCTATTGGACAAAGTTGGTCTAGACAGTACAAGGTGTGAAGACAC  
5,357 CGCCCTCTGCTGGAGAAGATGCTGGATTTTATTTACCTACAGGAAGAGACGTCTAAGT  
AGCAATTAGATGCTAACTAATGCTGCCTCAGGAAAGAATCAAAAGAGAAGAGTGAAAC  
CAGGCCGGGGCGGTGGCTCAGCCTATAATCTCAGCNACTTTGGGAGGCCAAGGCGAGG  
5,537 GGATCACGAGGTCAGGAGATCGAGACCTTCCTGGCTAACCCCGTCTCTATTAAGTACA  
AAAAATTAGCCGGGCATGGTGGCAGTGCCTGTAAATCCAGCTAGTCGGGAGGCTGAGGC  
AGGAGAATAGCTTAAACCCAGGTGGCGGAGGGTGCAGTGAGCTGAGATGTGCCACAGCAC  
5,717 TCCAGCCTGGGCAACAGAGCCAGACTCTATCTCAAAATAAGGAAAAATAAAAAAGAAAAG  
AAAGAAAGTCCATAAATGAGACTCCTAGAGATACTAAATGGTAGAATGGGAATTTGAAT  
TTAAATTTATAAGATGTTCACTCTCGGAGATCATAGGTCATTGTTGTCTCTCTCTTTTC  
5,897 ATGACAGGAAGTAAAGAGCTCTGACTATGTGCTAGGTAAGTACTCTGAGAACCTA  
ACATTTGATCTCCTTATTAAGTCTATTACTGCCCCATCTACAGATGAGAAAATGAGG  
CACAGGAAGTTAAGTTGGCAAGATCACACAGCCAGTAAGGGGCGAGCATTGAAGGTC  
6,077 ATTTTGCCTGCCTTATCCCCAGCCTCCAGGCAGTGGCAGAGTTAGCTCATTTTGGACAAA  
CAGCTCTCCAGACCAGACATTGTAAGCTATACTCAGGAATCATAGGAAAGATTATGATA  
GAATAATATATAGTTACAAAGAAAAGAAAATCCAATGGGAGAATATTTACTGTTTT  
6,257 CTATATTAAGTGTTTAATGTTTATGTTTTAGAGGAATATTGTTTATTATAGCAATTA  
GAAACAAAATGAGAAAAAAATCACCAGGATTCTACCTCCAGTTATTTTGTGATTT  
CCTTCCATTTTCCCCCATGTCTGTTTATATAATTGAACTATTATTCATGCAAGAGG  
6,437 TATTCTGATTTTCTCAGTTATTTTATTTTAAATTTGTAATAAACTTTTTCTT  
CTGAGACAGTCTCGCTGTGTCACCCAGGCTGAAGTGACGCGTGAATCTTGGCTCACTG  
TGACCTCCAGTCTCAAGCAATCTTCCTGCCTCAGCCTCCTAAGTAGCTGGGACTGCAGA  
6,617 TGTGTGCTACACATCCAGCTAATTTTAAAAATCTTTTTCTCTTTTTGGTAGAAATGG  
GAGTCTCTATGTTGCCAGCCTGTATTGAATCCTGGGCTCAAGTGATCCTCCACCT  
CAGTCTCCAAAGTCTGGGATTTCAAGCATGAGCTACCACACCCAACTAATTTTATT  
6,797 TTTATTTATTAAAAAAAAGTTTTGCTACCCGCTCTCACCCTCTCACTCATTTTT  
AAGTATAGGTTTTCTACAATGCCACATTGCTTTTCCATAAAAAGTACTTTCGGCCGGG  
TGTGGTGGCTCACACCTGTAATCCAGCACCTTTGGGAGACTGAGGCGGGTGATCACCTG

FIG.4E

SUBSTITUTE SHEET

10/25

6,977 AGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCATCCCTACTAAAAATA  
CAAAAATTAGCTGGATATGATTGTGGGCACCTGTCATCCAGCTACTCGGGAGGCTGAGG  
CAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATTGTGCCACTGC  
7,157 ATCCCAGCCAGGGCAACAGAGCGAGACTTCATCTCAAAAAAAAAAAAAAAAAAGTACTTTT  
CTTCATTTGGTTAGTATTCTCTTATGAGTTGATGCCCTTGTAATTTATCTGAATGTTTTCC  
ATTATTTTGTGGTGAGCTTTAAACTACCCCTTCCTGACTTTCAAGAATCCTAGACATGCT  
7,337 CCTTGTTGCTAGGTAATTATTAGTGCACCTATTAGAATAAGTATATGCTTGGAGTGGG  
GAGGAGATGAACTTTTGAAGGGCGGTGAAGTATTTCTACCACCAGGCCTTTGTCTTTG  
CTAAACTGAGGAAGGAAGATTTTATTTTATTAGCTAACAAAGAACCTCCTATATAGGCCG  
7,517 GGCATGGTGGCTCAGCCTGTAATCCTCACATTTTCAGAGGCCAAGGTGGGTGGATTGCC  
TGAGCTCAGGAGTTTGAGACCAGCTTGGGCAACATGGCAAAACCCCATCTCTACTAAAAA  
TACTAAAAATTAGCTGGGCGTGGTGGTGAGTGCCCTGTAATCCAGTTACTCCAGAGGGTG  
7,697 AGGCAGAAAATTGCTTGAACCCGGGAGGTGGAGGTTGCCATGAGCCGAGATCGTGACAGT  
GCACTCCAGCCTGGGGCAGACAGCAAGACTCTCTCTCAATAAAATAAAATAAAATAAAAT  
AAAATAAATACATAAATAAATCTCCTATATAACCTCATAATATCAGATTTGGAGCCTTTT  
7,877 CCATAGAAATGAAATTCAGAAGAAGCTGAGACTCAGATATTCCAAGCTGCCTGGTGTCT  
GTGAATAGAGGAGACTTGTCTTGTGAAATCTGAGTGCAAAGACACAGGACAAATTGTTA  
TCTACTTTTCATTCTAAGGATACTGTATGGCCCTAAACACAAAGAACTAGAATTCTGTG  
8,057 ATACCACGGGTACTCCACAGTGTGTTCCCTCCCTTTCTGAACCTGATTTGTCTCATCTC  
TATGAAAAGATGTGGGCTTTGGGGTCAGATGTGGGTTGGAATCCTAGCGCTGTGTGGCT  
GCAATTTTCTTTGTGTAATAATTGAGATAATAGTACAAAAGTAACAACAGTTAATATTAT  
8,237 CAAGTGCTTACTGTGTGCTGGCACTGTGTTAAATCTCTAAGTGATTTTCTCATTTAA  
TTTTTGTGATAGGCTTATGACACTATTAGTATCTTCATATTACAGTGAGGGTTCAGAGAA  
GTTAAGGTTCCATAACTAGTCAGCAGACCTGGGACTTCACTCCAGGCAGCTGATTCCAAA  
8,417 GCCTATTCTAACTTTAACTGCTACTTTTTGGAGTGTGTAGGAGGACAATTTATATAA  
AATGTTGGCACATAGTGGGTGCTGCTGTTATATGAATGGGCACAAAATCTGTCTACATTT  
TGCCTTTTACCAAAATTTAGAATCTATTTAGTTAAACCTTCTTAGGGCGGGTGGAGTGCA  
8,597 GTTGCTCATTCTGTAAATCTCAGCACACTGGGAGGCCAAGGCAGGAGGATTGCTTGAGCC  
CAGGTGTTTGAGACCAGCCTGGGCACATAGTGAGACCCCATCTCTCCAAAAACAACA  
AACAAACAAAAACAAAAACAACTAGCTGGGCGTTGTGGTGCCCTGTATTCCCAGCTAC  
8,777 TCAAGAGGCTGGGGTGGGAGAATGGCTTGAGCCAGGAGTTCAAGGTTGCAGTGAGCTAT  
GATCACAGTACTGCACTCCAGCTTGGGCAGCCGACTGAGACCTGTCTCGAAAAAAAAT  
AAAAATAAAACTTCTTAGGACAGAGTGATTAGAAGCTCTCTAGTAGATACTTAGTAACA  
8,957 ATGTGGGTTCCTCGGGCAG

FIG.4F

SUBSTITUTE SHEET

11/25

GTTATCCTACGCAGAGAGCCAGGTACCAATGGGTGCGCTGC  
AATCCAGACAGTAATTCTGCAAACTGCCTTGAAGAAAAAGGACCAATGTTCCGAACACTTT EXON 2  
CCAGGTGAATCCAACAAGATCCCCGTCTGAGGACTGACCTTTTTCC

GTAAAGTGGACTTT  
9,137 TCTCTAATTAATTAATTAATTACTTATTTATTTGAGACGGAGTTTCACTTTTCTTGCCCA  
GGCTGGAGTGCAATGGCGCAATCTTAGCTCACTGCAACCTCCGCCTCTGGGTTCAAGCG  
ATTCTCCTGCTTCAGCCTCTGGAGCAGCTGGGATTTGAGGCGCCTGCCACCATGCCAGC  
9,317 TAATTTTTTTTTTTTTTTTTTTTTTGGAGACGGAGTCTCACTCTGTTGCTCAGGCTGGAGTG  
CAGTGGCGCAATCTCGGCTCACTGCAAGCTCCACCTCCTGGATTACGCCATTCTCCCGC  
CTCAGCCTCCCGAGTAGCTGGGACTACAGGCACCCGCCACCGCCCGGCTAATTTTTTT  
9,497 GTATTTTAGTAGAGACGGGGTTTACCTTATTAGCCAGGATGGTCTCGATCTCCTGACC  
TAGTGATCCGCCCGCCTTGACCTCCCAAATGCTGGGATTACAGGCGTGAGCCACTGCGC  
CTGGCCTAATTTTTGTATTATTAGTAGAGACGGGCTTTCATCATCTTGCCAGGCTGGT  
9,677 CTCAAATCCTGACCTCAGGTGATCCACCCACCTTGGCCTCCCAAAGTGTGGGATTACA  
AGCATGAGCCACTGTACCCGGCCTTTTCTCTAATTTTAAAGTGTCTGTAAATTCACAACC  
TCTTGGCAGAGATGTGGGAGTGTTTTCTTCAAGCTGTCCAGAGTGTTTGCTTCGAGCT  
9,857 CTTGCTTTGGTAGTTTGGCTTACTCTGCAGTACATGGTAAAGTGTACTGTATATACT  
GGCATATGACATGTGCGAGTATACATGATTCACCTATGTTTTTGAATTTTTTTTGTGGA  
TGGTAGAGAGGAGCATTGAGGACTTTTCATCAACAGGTATTGAAAATGATTGAACATTGT  
10,037 TTTATTTGTGTAACAGAACACACTATATATAAAAAATCCAATAATTAACGAATGGATAA  
GCAAAATGTGGTATAAGCATACAAAGGAATATTATTGGTCAAAAAAGAAATGAAGTACT  
GATACATGCTACAACATAGATAAACCTTGGAAACATTATGCAGAGCGAAGGAAGGCCAGA  
10,217 CACCAAAAGCCACATATTGTATGATTCCATTAGATGAAATGTCCAGAATAGGCAATCC  
CTAGAGGCAGAAAGTAGATTAGTGGGTACAGGGGCTGGGAAAGGGAGGAATAAGGAGT  
GACTGCTAATGGGTATGAGGGTTTTTTTTGGAGGAGGTGATTAAATGTTCTTCTGCCAG  
10,397 GTGTGGTGGCTCATGCCGTGAATCCAGCACTTTGGAGGCGGAGGCGGGAGGATTGTTT  
GAGCCAGGAGTTTGAGGCCAGCCTGGGCAACATAGTGAGACCTATCTCTATTTCAAAT  
ACATTTTTATATTAATAAATGTTCTTCAAGTAGTTGGTAATTATTTTAAAAATGGCC  
10,577 AGGTGCAGAGGCTCATGCCGTGAATCCAGCACCTTGGGAGGCTGAGGTGGGAGGATCCC  
TTGAGCCCAGGAGGTTTGGGACCAGCCTGGGCCATACAGCAAGACCCTGTCTCTACAAAA  
AATACAAAAATTAGCTAGGCATAGTGATGTGCACCTGTGGTCCAGCTACTCGGAGGCT

FIG.4G

SUBSTITUTE SHEET

12/25

10,757 GAGGTAGGAGAATCTCTTGAGCCTATGTTGAGCCTGCAGTGAACCCATTTATGCCATTG  
CACTCCAGCCTAGGCAACAGAGTAAGACTCTGTCTCAAAAAAGAAAAAAATTAGGG  
AAAGGAAGAATAATTAGCCAAGACTTGTAACCAAAAATCAATCTCTCTTTTGATCAC  
10,937 ATAAACTTGCTTTAACTTGCAAAAAAGACCTGATATAAATTCATAAGTAACAAAAAT  
TGAATTATATTAGAAACCATTAATTCATGAATACTAAAGCTATGTAGGATGTAGCAAAA  
TATACATATTAAGAAAAGGATTATCATAAAGTTTTAATCTCCAGGCTCAACCTAGAAA  
11,117 ATCACTCTCCTCAAAGCCAGGGTTAATCATCATGCTCCAAACCAGGTACATTTACATCA  
CTTTGGGATCCTGGCAACTTTCTCTTTTGTCTTTTTTTTTTTTGGAGACAGGGTCTCCT  
CTGTCAACCCAGGATGGAGTGCAGTGGTGTGATCATAGCTCACTGCAGCCTGCAACTCCTC  
11,297 AAGTGATCTTCCTGCCCTCAGCCTCCCAAGTAGCGGACCACAGGCACACAGCACCATGCC  
CATCTAATTAAAAAATTTTTTTTGTAGAGACAGGGGTCTCTGTACATTTCCAGGCTG  
GTCTACTCTTAAGCTCAAGCAGTCTCCACCTCAGCCTCCCAAGTGCCGGAATTA  
11,477 CAGTCATGAGCCACCATTCAGCGCTGGTGACTTTCTCCATCACTGGTGACTTTCTCCA  
TCACTGGTATTCACTGCATTAGTGATGACATCATTACAATCTTCAATATGCAACTTTGTA  
GTCCTACTCTTGCACTTCTACTTTAAAGCCTCAGCATTAAAGTTGAATGTAATTACAG  
11,657 CATCTTCATTACTTTAAATCATTGGTTTCAATAGTAATTCATTTAAATCTAAATGTTA  
GGCTGCAGTGGCTCATGCTGTAAATCCCCCAGTTTGGGAGACTGAGGTGGGAGAATCAC  
TTGAGGCCAAGAATTTGAGACCAGCCTGGGCAACACGGCAAGACCCCATCTCTAAAAATT  
11,837 AGTGGCCCGGCGCTGTGCCTCAGCCTGTAAATCCCAACACTTTGGGAGGCGGAGGCGGA  
TAGCTTGAGGTGAGGAGTTCAAGATCAGCCTGGCCAACATGGCGGAAACCCATTTCTACT  
AAAAATACAAAAATTAGCTGGGCATGGTGGCAGCCTGTAAATCCAGCTATTGAGAGGCC  
12,017 GAGGCAGGCAGACTGGGAGGCCAAGGCAGGCAGATTGCTTTGAGACCTGCTGGGTAACA  
TGGAGAAATCCTGTCTCTACAGAAAAATACAAAAATTAGCCAAGCATGGAGAAACCTCGT  
CTCTATAGAAAGACACAAAACTAGCCATGCATGCTGTGGTCCAGCTACTCGAAAGGCT  
12,197 GAGATGGGAGGATTGCTTGATCCTGAGAGGTCAAGGCTGAAGTGAGCCATGGTGTGGCAC  
TGCACCTCCAGCCAGGCTGACAGACTAAACCTTGCTCTAAAAATAAATAACACATTTAAA  
ATAAATAAATAACAATTAACCTAAATTAATAAATAAATAAATAAATGTTAAGAGAATAGCT  
12,377 CAAATTCTCCAAAGAACTCTTGACACCATTCCTCCTCTCTCAAATCTCTATTTTCCT  
TCCCCAAAGCCAGTAACCTGCTCTCACCCTGACCTGTGCTTTCTTTCCGTCATTGCGA  
AAGAATGGTCTTGCTTCTGTGCTGATCCCAACCCTTTGGCCCTCAGATCCTCCTGTCC  
12,557 TTCCCTGGCCCTGCTCTGTATTGGCTGTGGGGTGGGGTGGCGGTGGAACAGACCCCTGG  
GGTCTGCATTTCTCAGGCTCCAGGGCTGTGGCTGACTTTGGCCAATGGGAGGCAAGGAC  
GGGAGACTGAGAGCTTGGGAGGAAGGGAGAGAGGTATGTTTCTCTCCTTACTCCCTGCC

FIG.4H

SUBSTITUTE SHEET



12,737 TGGGGTGGCACCCTGGGGCAGGACTCTGTTTTGCCCATGGCCTCAGCTCCCACCAGATGCC  
TCTAGTCCCTGGGCTCAGGAAATAGACAACCTCCTCCACTATCGCTGTAGCCCAAGGAG  
GGAAGTATTTTTTCTTTCTTTCTTTCTTTCTTTCTTTTACAGAGTCTCACTCT  
12,917 TGTGGCCAGGCTGGAGTGCAGTGGTGGATCTCAGCTCACTGCCACCTCTGCCTCCCG  
GTTCAAGTGATTCTCCTGCCTCAGCCTCCAGGGTGTGCCACTATGCCCAGCTAATTTTTG  
TATTTTTGGTAGAGACGGGGTTTTGCCATGTTGACCAGGCTGGTCTTGAACCTCTGACCC  
13,097 CAAATGATCTGCCTGCCTCGTCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGTG  
CCTGGCCGAAGGAAATATTTCTTGCTATTGCTAATCTCTGGGTACCTCGCTATCCCC  
ATTTAGCTTCACTTCTCCTCCATCACCTGTATGAGGAATCCCTCTGTGTTAAATATCTG  
13,277 GAGAAGTTTCTGATTGGACCTGGCTGTTGCAGCTTCCAAGGCCACCTCTCTTTGTGGC  
TGGTATCCTTTTCCATGCATCTTCTCCAGGACTTCCATTCTGCAGTTATCTCTGAAC  
TCAGTGTCTTCTTCCATCAGTATAGGGGTGGACTTTAGTATCTCCTATGTTTAGGCAAC  
13,457 ATCTCTCCTTTGACTCTGGCTCTTCTCCAGTGGTGGCCTTCTCTGCTCCTCTTCAAT  
AACACCTCCTGAAAGGGCCACCATGCCGTGCCCCCTCCTTTCTCACCCTCTGTGGCT  
GGACTTCTGTTCTTCACTCCACCTGGTGGACAAAGTCACTGATTACTTCTCTATTTTC  
13,637 AGCTTACTTGATCCTTAATTGCCTTCAAAAACAGCTAACTGGGCCATGCATGTAATCCA  
GCACCTCGGGAGGCCAAGGCAGGAGGATCACTTGAGCCAGGAGTTCAGGACCAGCCTGC  
CTGGGCAACATAGTGAGACCTATCTACAAAAATAGAAAAATTAGCCGGCGTTGTGAC  
13,817 TCATGCTTGTGGTCCAGCTACAAAGGAAGCTGAGGTGGGAGGATGGCTTGAGTCCGGGA  
GGGTGAGGCTGCAGTGAGCCATGATCAGCCACTGCACCTCAGCCTGCACAACAGAAGGA  
GGCCCTTTCTGTAAAAAATGGTTGACCACTCCTTCTTGAATGCTTTTTTCTTG  
13,997 AGGCTTCCATGCCCTGCCTTATCCTGTTTCTTCTTCTTCTGCTTGTGCTTTTCTC  
TGCTCAGTATTTAACATGTTGGTGTGACCTGGCTCTGGCTGGGCCCTTCTCTATCT  
ACGTGCTTCTCTCGACGACCTCCATCGGTTGCATGGTTTAACTACCAATCTGTGATT  
14,177 CTAGCTCCGACACCCAGGCTAAAGTAGCCACCTGGTCACTCCCTATTACATTGGTCAAT  
TTCATTTCTCTGTGCCACCTATGATTTTCTGATTTATTTATTCACTTTTCATTGTCTGT  
CTTCCCCACTAAAAATAAACTTCTTGAGAAGGGGCTTCATCGATCTGCCTCTGTTCTAT  
14,357 CCCAGGCCCTCAAAACAAGGACCAGATATTCAACAAATATTTATTGAATGCGTACATGAA  
TAAAACTCTAATTGGTTGTATGCTGGTGGTTATTATTTTCATGGAGGAAATGACTTGT  
AGGCTGTGACACTCAGCTTTTGTCTCTGATGCTTTGTTGCCCTGTTCTGTACCGAGGGC  
14,537 TGTGCTCATTGCTCTGCCATTTTGTCTCTTGAATTTCTAATCATCAGCTCAACCCA  
GAAGGCAGCCTTACCTTTCAAGCACTTTCAGCTGAATGAGTGCAAGTTGGAGGCAGGGTC  
ATTTTTGTAGGAAATGAATGTTATATGCTGGTAAATATAAAGCTTAGCTTTTTACA

FIG.4I

SUBSTITUTE SHEET

14/25

14,717 AAGAATTTCTCAAAGTGAGCTTTGTTGAAGCCCTGTAATTGTTAGAACTTTTATGGAA  
ATTTTAATTTAGGAAAAATGTCATCTGTTTGGGCTGACTTAGTTGTTAGTTGTTTGCC  
TTTCTTTTTTTTGGTGGAGGGTATGGAGTTTGTCTTGTAAACCCAGGCTGGAGTGCAGT  
14,897 GGC CGCATCTCGGCTCACTGCAACCTCCGCTCCTGGGTTCAAGCGATTCTCTCACCTCA  
GCCTTCCGAGTAGCTGGGATTACAGGCATGCACCACCACACTTGGCTAATTTTGTATTT  
TAAGTAGAGACCGGGTTTCACTATGTTGGTCAGGCTGGTTTCGAACCTCTGACCTCAAGT  
15,077 GATCACCACCTTGGCTCCCAAAGTGCTTGGATTACAGACATGAGCCACCACACCCGGC  
CAAGAGGACTTCTTTAAAAATGATTTCTTGGGCCGGGTGCAGTGGCTCACACCTGTAAT  
CCCAGCACTTTGGGAGGCTGAGGTGGGTGGTTCACAAGGTCAGGAGTTTGAGATCAGCCT  
15,257 GGCCAATATGGTGAACTCCATCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGC  
GCACCCCTGTTGTCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGA  
GGTGGAGGTTGCAGTGAGCCGAGATGGCACCCTGCACTCCAGCCTGGGCAACAGAGCAA  
15,437 GACTCTGCCTCCAAAAATAAAATTAATGATTTCTTAAGTAAATTTCAAATATAGAAT  
GTATATGCTAGTGATAACAAATTAACACTGTTTATGCAAGTCTGCAATAGGTAGATGTG  
AAGTTGATAGGTGCAATAAGTATAGGCAACACATAGGAACATTGACCTGTTTTTTGT  
15,617 TGATTTTAAACATTGAATAATTGGGAAGCTTTTAAATCTCTTAATTTGAGCAACTAGAT  
GGCTGTATTTATCTCCTTATATTAAAAACTATTATAATTATCTTTCCACATATCAA  
CTCCACTGGTTTTTTCCATTTTCTTTCATACTTCAG

15,797 GAATCGTATCTTCCCACTTTCTGAGGACTACTCTGGATCAGGCTTCGGCTCCGGCTCCGG  
CTCTGGATCAGGATCTGGGAGTGGCTTCTAACGGAAATGGAACAGGATTACCAACTAGT  
AGACGAAAGTGATGCTTCCATGACAACCTTAGGTCTCTTGACAGGAATCTGCCCTCAGA  
15,977 CAGCCAGGACTTGGGTCAACATGGATTAGAAGAGGATTTTATGTTATAAAGAGGATTT  
CCCACCTTGACACCAGGCAATGTAGTTAGCATATTTTATGTACCATGGTTATATGATTAA  
TCTTGGGACAAAGATTTTATAGAAATTTTAAACATCTGAAAAGAAGCTTAAGTTTA  
16,157 TCATCCTTTTTTTCTCATGAATCTTAAAGGATTATGCTTAAATGCTGTTATCTATCTT  
ATTGTTCTTGAAATACCTGCATTTTTTGGTATCATGTTCAACCAACATCATTATGAAAT  
TAATTAGATTCCCATGGCCATAAAATGGCTTTAAAGAATATATATATTTTTAAAGTAG  
16,337 CTTGAGAAAGCAAATTGGCAGGTAATATTTCATACCTAAATTAAGACTCTGACTTGGATTG  
TGAATTATAATGATATGCCCCTTTCTTATAAAAAACAAAAAAATAATGAACACAGT  
GAATTTGTAGAGTGGGGTATTTGACATATTTACAGGGTGGAGTGTACTATATACTATT  
16,517 ACCTTTGAATGTGTTTGAGAGCTAGTGGATGTGTTGTCTACAAGTATGATTGCTGTTA  
CATAACACCCCAAATTAACCTCCAAATTAACACAGTTGTGCTGTCAATACCTCATACT  
GCTTTACCTTTTTTCTCGGATATCTGTGATTTTCAAATGTTACTATATATTAAGCAG  
16,697 AAATATAACC

AAAGACGAGAATCCAGGACTT

EXON 3

FIG.4J

SUBSTITUTE SHEET

-504 AATTCTAGCAGACTCTGGACGTTAACGGAGACCGCTCATCCTGGGGGCTGAGAACCAGCTCGGCTCGGAATGTT  
 -429 CCCTGCTTGTGCGTGACTCTGTGCGCGCCAGCTTCTCTTTGATGTGCGCTGTGGATGAGCCGAGCTCAGTTCTG  
 -354 GAACAGCTGAGTCCCTCTGTCTGTTAGATTGTTACCTGAAGGAAGGGAGGGGGAAGAAAGTCTGATTCTGACTT  
 -279 TTTGATGGGGAAAACCTTTTTTTTAAACATGCAATGACAGATGGCAGAGCTTTTTGGAAAAAGAAAAATAATA  
 -204 ACCACACAGCAACGCCCTAGGGGGAGTCCGGTGGAGTTTCATCATGGGTATGAACAGTTGTTGTTTTTCAACT  
 -129 TTCTTCTCTTCTGGGTGTTGATGTGGATCTCTTTCTATTTGTTCAAGAAACTGTGACGTGTGTTCTTGGGCAG

-54 GTGCTGAGGTTTTGGAACCTCTTTCTAAAGGGACAGAAAGAGCACCTGCTACATTGCTAATCCAGAGGCTGA

START

1

GTGGAGCCGAGCTGGTCAGG

ATG	CAG	GTT	CCC	GTC	GGC	AGC	AGG	CTT	GTC	CTG	GCT	CTC
MET	GLN	VAL	PRO	VAL	GLY	SER	ARG	LEU	VAL	LEU	ALA	LEU

exon 1-|

GCC	TTC	GTC	CTG	GTT	TGG	GGA	TCT	TCA	GTG	CAA	Ggt	aaagagaccaggaattttaattc-
ALA	PHE	VAL	LEU	VAL	TRP	GLY	SER	SER	VAL	GLN(GLY)		

|—exon 2

-( 8 kb)- ggttccttggttcgaca gGT TAT CCT GCT CGG AGA GCC AGG TAC CAG TGG GTC  
 (GLY) TYR PRO ALA ARG ARG ALA ARG TYR GLN TRP VAL

CGC TGC AAA CCG AAT GGC TTT TTT GCG AAC TGC ATC GAG GAG AAG GGA CCA CAG TTT  
 ARG CYS LYS PRO ASN GLY PHE PHE ALA ASN CYS ILE GLU GLU LYS GLY PRO GLN PHE

exon 2—|

GAC CTA ATA GAT GAA TCC AAT AAC ATC GGC CCT CCC ATG AAT AAT CCT GTT TTg too  
 ASP LEU ILE ASP GLU SER ASN ASN ILE GLY PRO PRO MET ASN ASN PRO VAL(LEU)

|—exon 3

gtagactttcatgat -( 4 kb)-tttttctttgtatttt agG ATG GAA GGA CCC TCA AAA GAT  
 (LEU)MET GLU GLY PRO SER LYS ASP

FIG.5A

TTC ATC TCC AAT TAT	GAT GAC	TAT	GGG TCA GGT TCG GGC TCC GGC TCT GGC TCC GGC
PHE ILE SER ASN TYR	ASP ASP	TYR	GLY SER GLY SER GLY SER GLY SER GLY SER GLY

TCT GGC TCG GGT TCC GGC TCC GGA AGT GGC	TTC CTA GGT GAC ATG GAA TCC GAA TAC
SER GLY SER GLY SER GLY SER GLY SER GLY	PHE LEU GLY ASP MET GLU TRP GLU TYR

CAG CCA ACA GAT GAA AGC AAT ATT GTC TAT TTC AAC TAT AAG CCT TTT GAC AGG ATT  
GLN PRO THR ASP GLU SER ASN ILE VAL TYR PHE ASN TYR LYS PRO PHE ASP ARG ILE

CTC ACT GAG CAA AAC CAA GAC CAA CCA GAA GAC GAT TTT ATT ATA TGA  
LEU THR GLU GLN ASN GLN ASP GLN PRO GLU ASP ASP PHE ILE ILE STOP

ATGTGACGGTCTCTGTCTCCACCTCCATGTGGAACAATGTATTTCAGTATACTTAGTGTACCACGTTTAAATGA  
CCAGTCTCAGGATAAAGAGTTTTACAGAAAATTTAAATGCCTGGAAAAGACTCTTGAATCCTGTTACCCCTTTC  
CTCATTAACCTCGTAAGGAATTATGCTTTAATGCTGTTACCTATCTTGTGTTCTGGAAAATGCCTGCATTTATGT  
GTATTGAATCAACATTTAAGAAATTAACACACACCCCATTTATATACAATAACTTTCAAAGCCATACTGGTTTT  
GAAAATTTTAATTTGATAGCAAGTTGATGAACAATCTTTCATACCTAAAGTGTTCAGGAACCCAACCTCGATTGT  
GAATTACAAATATATTCCTTTATGTGATTAAAAAGAAAATAAAGTG

FIG.5B

17/25

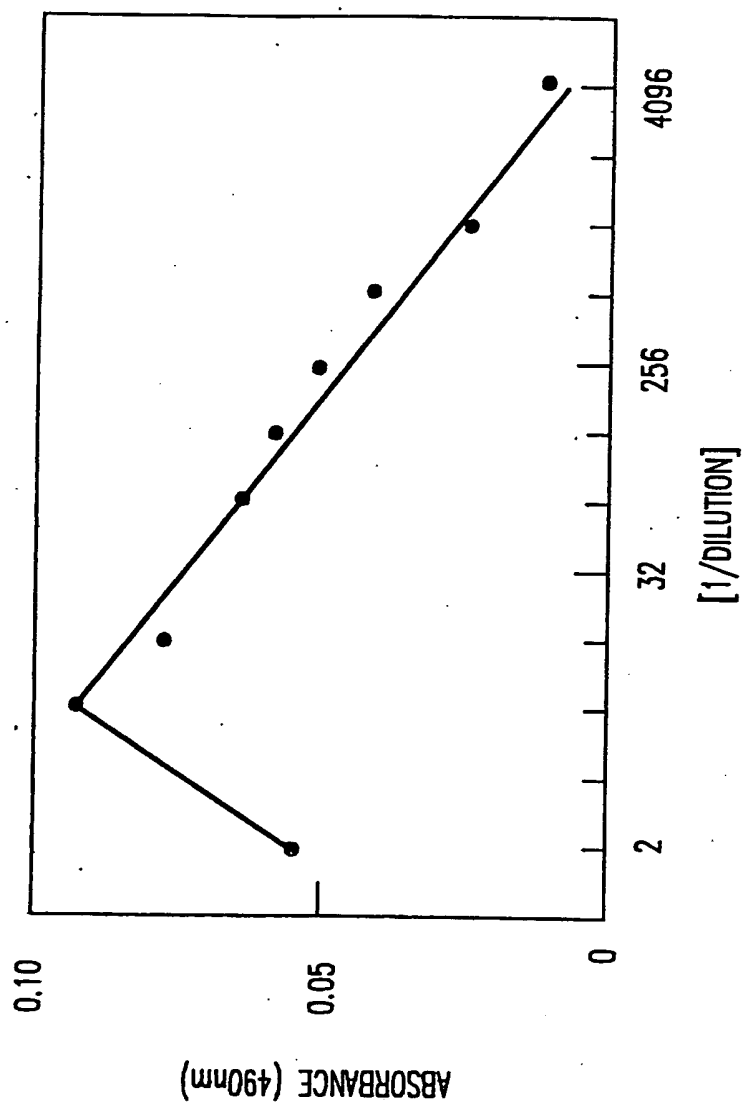


FIG.6

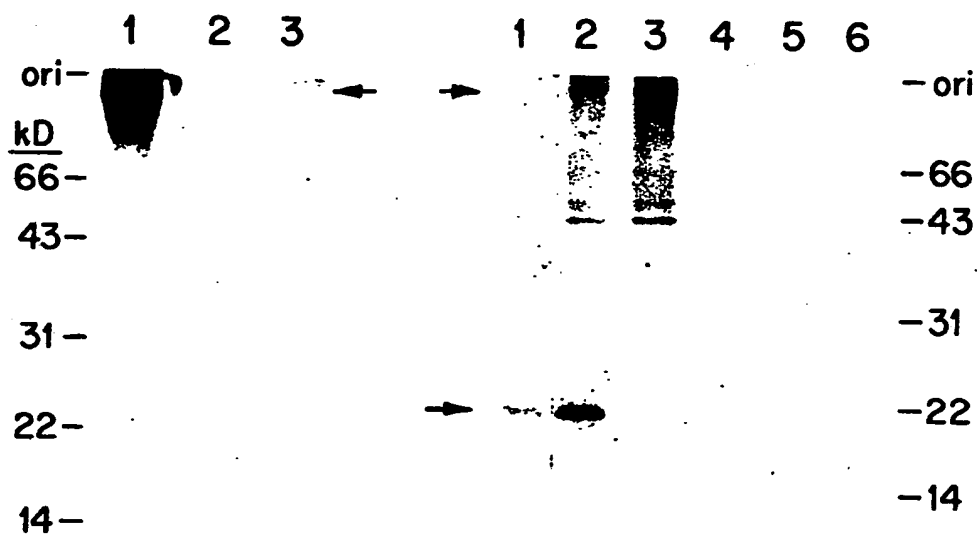












FIG. 7A

FIG. 7B

19/25

CONSTRUCT		RELATIVE hGH EXPRESSION	
		RBL-1 CELL	Rat-1 FIBROBLAST
pPG(-504/+24)hGH	-504 	$0.89 \pm 0.15$ (18)	$0.05 \pm 0.02$ (18)
pPG(-423/+24)hGH	-423 	$0.82 \pm 0.13$ (6)	ND
pPG(-333/+24)hGH	-333 	$0.79 \pm 0.11$ (6)	$0.03 \pm 0.01$ (6)
pPG(-250/+24)hGH	-250 	$1.24 \pm 0.23$ (8)	$0.05 \pm 0.02$ (8)
pPG(-190/+24)hGH	-190 	$2.52 \pm 0.49$ (8)	$1.06 \pm 0.27$ (8)
pPG(-118/+24)hGH	-118 	$3.21 \pm 1.61$ (14)	$1.21 \pm 0.65$ (14)
pPG(-81/+24)hGH	-81 	$0.46 \pm 0.16$ (10)	$0.22 \pm 0.12$ (10)
pPG(-63/+24)hGH	-63 	$0.34 \pm 0.12$ (8)	$0.20 \pm 0.10$ (8)
pPG(-40/+24)hGH	-40 	$0.41 \pm 0.12$ (8)	$0.32 \pm 0.09$ (8)
pPG(-20/+24)hGH	-20 	0.00 (8)	0.00 (8)

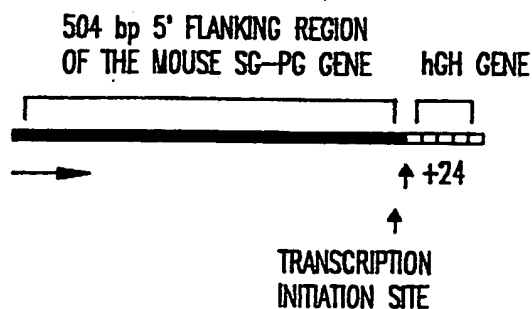


FIG.8A

SUBSTITUTE SHEET

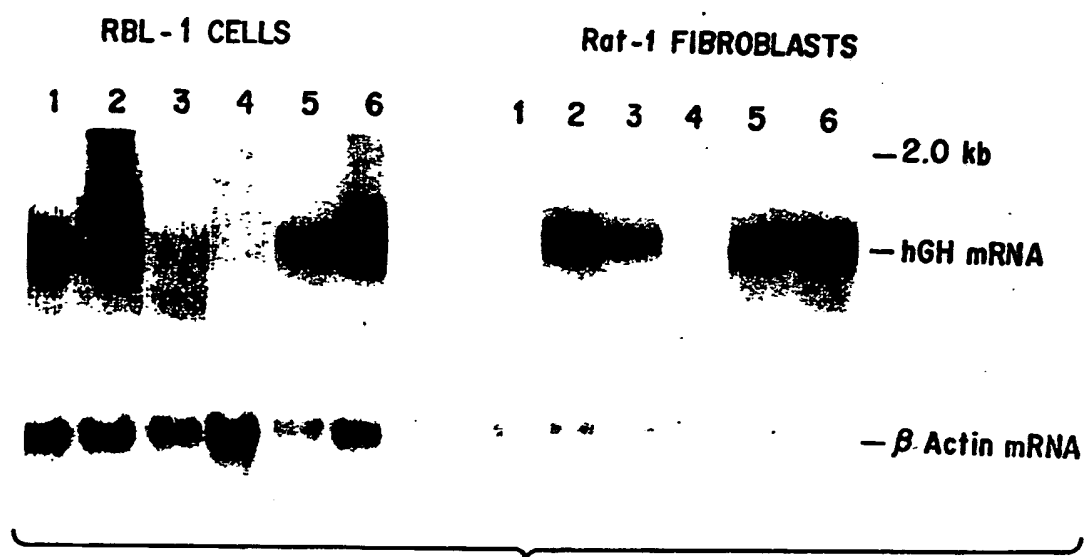


FIG.8B



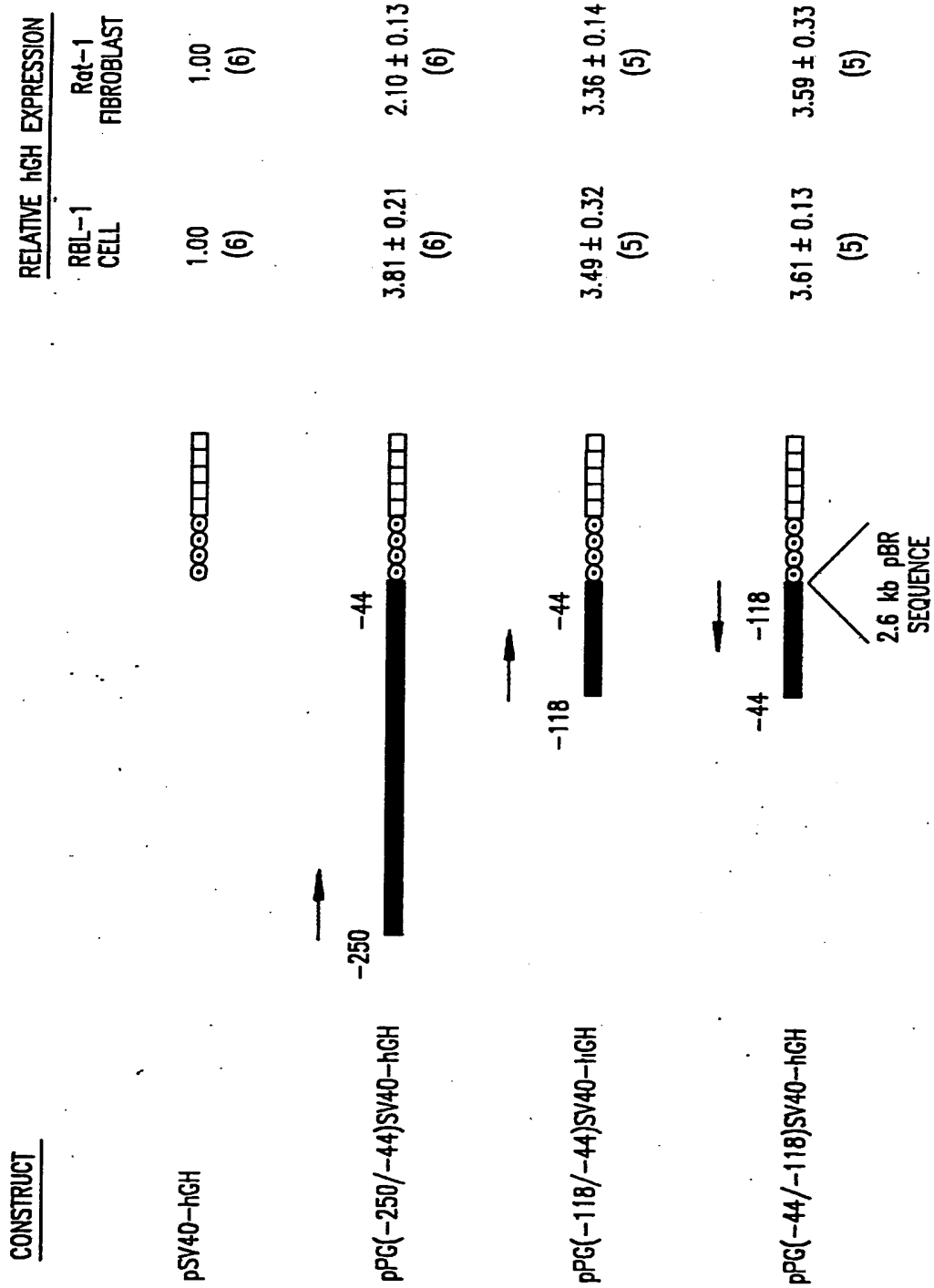


FIG.9

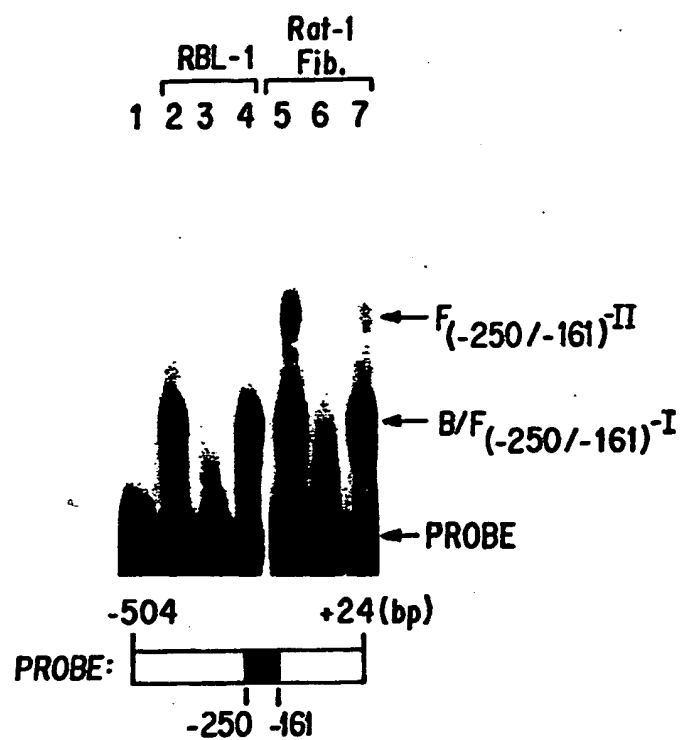


FIG.10

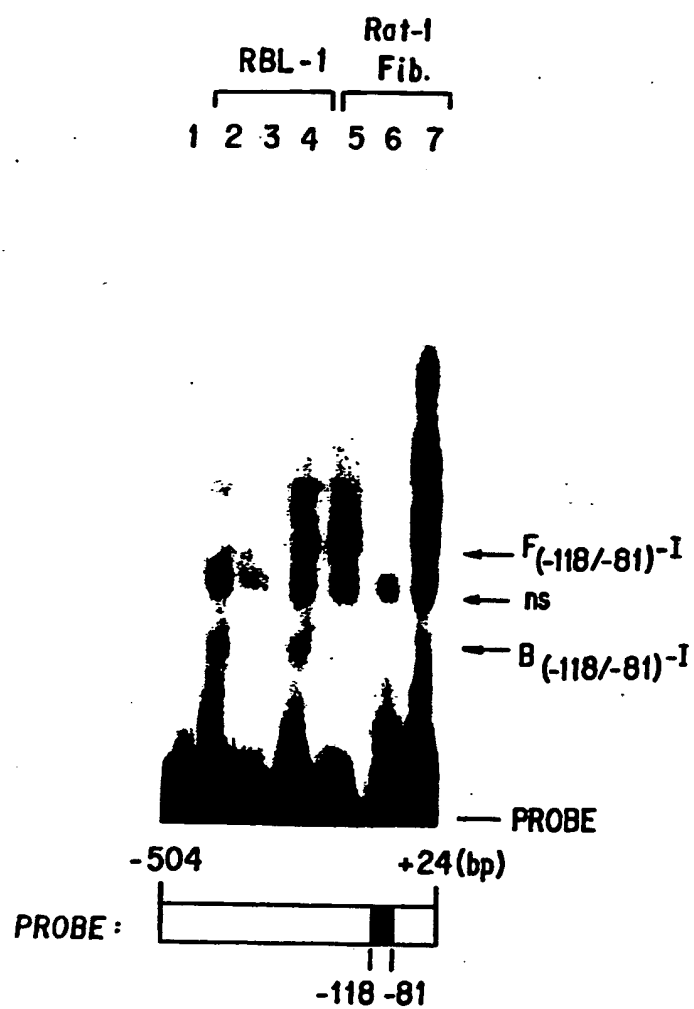


FIG. 11

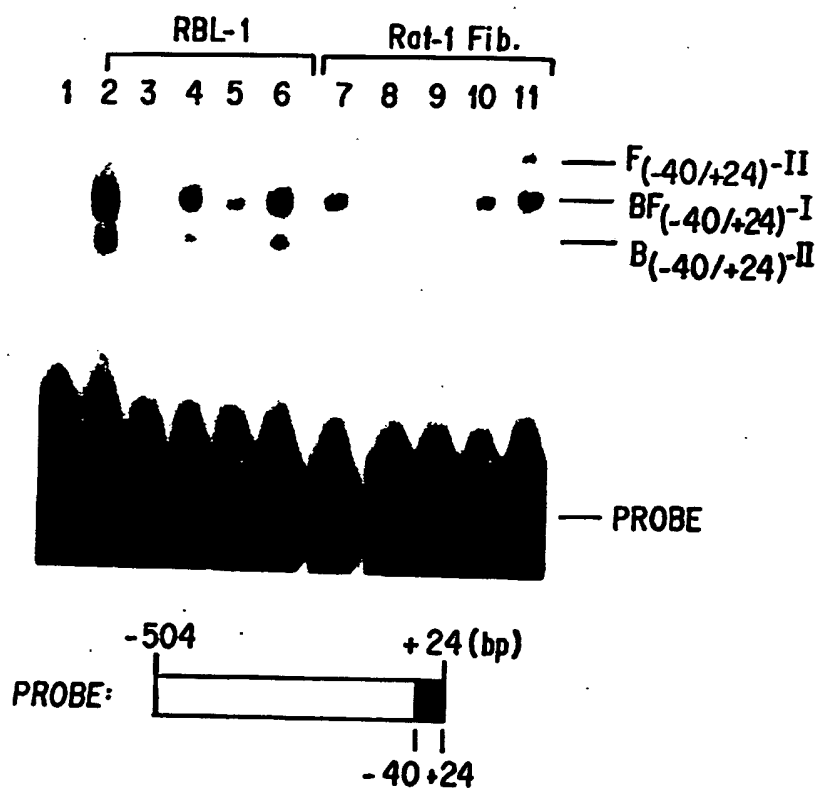


FIG. 12

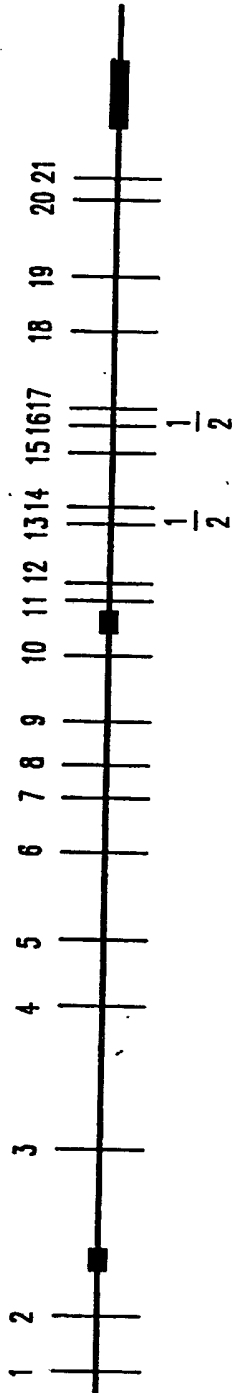


FIG.13A

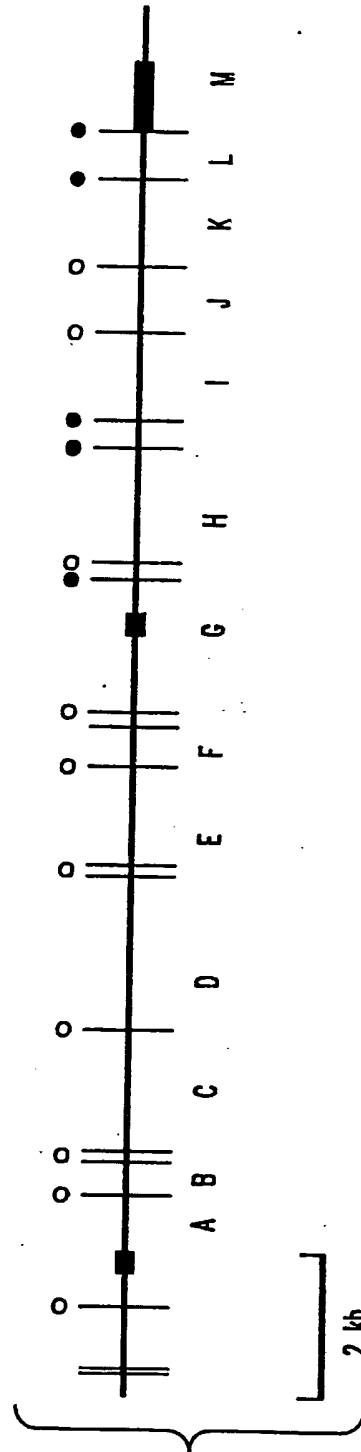


FIG.13B

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US92/11194

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : Please See Extra Sheet.

US CL : 435/69.1, 172.3, 252.3, 240.1, 240.2, 320.1; 536/24.1

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 172.3, 252.3, 240.1, 240.2, 320.1; 536/24.1; 935/6, 36, 66, 70, 72

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Journal of Biological Chemistry, Volume 264, No. 28, issued 05 October 1989. S. Avraham et al., "Cloning and Characterization of the Mouse Gene That Encodes the Peptide Core of Secretory Granule Proteoglycans and Expression of This Gene in Transfected Rat-1 Fibroblasts", pages 16719-16726, especially figure 2.	1-30
Y	Journal of Biological Chemistry, Volume 265, No. 10, issued 05 April 1990, C.F. Nicodemus et al. "Characterization of the Human Gene Taht Encodes the Peptide Core of Secretory Granule Proteoglycans in Promyelocytic Leukemia HL-60 Cells and Analysis of the Translated Product", pages 5889-5896, especially figure 4.	1-30

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be part of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier document published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"A" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

02 March 1993

Date of mailing of the international search report

31 MAR 1993

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Authorized officer

GABRIELE E. BUGAISKY

Facsimile No. NOT APPLICABLE

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US92/11194

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Gene, Volume 93, issued 1990, T. Angerth et al., "Cloning and structural analysis of a gene encoding a mouse mastocytoma proteoglycan core protein; analysis of its evolutionary relation to three cross hybridizing regions in the mouse genome", pages 235-240, especially figure 2A.	1-30
Y	<sup>A</sup> WO 90/00606 (Stevens et al.) 25 January 1990, Figure 4.	1-30
X	Annual Review of Biochemistry, Volume 58, published 1989, P.F. Johnson et al., "Eukaryotic transcriptional regulatory proteins", pages 799-839, especially Table I.	31-37
X	Nucleic Acids Research, Volume 16, No. 5, issued 25 March 1989, E. Wingender, "Compilation of transcription regulating proteins", pages 1879-1902, entire document.	31-37
Y	Cell, Volume 46, issued 12 September 1986, S. McKnight et al., "Transcriptional Selectivity of Viral Genes in Mammalian Cells", pages 795-805, entire document.	1-30
Y	US, A, 4,663,281 (Gillies et al.) 05 May 1987, entire document.	21-29
Y	Molecular and Cellular Biology, Volume 5, No. 9, issued September 1985, L.-J. Chang et al., "Gene Expression from both intronless and intron-containing Rous Sarcoma Virus clones is specifically inhibited by anti-sense RNA", pages 2341-2348, entire document.	30

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US92/11194

## A. CLASSIFICATION OF SUBJECT MATTER:

IPC (5):

C07H 21/00; C07K 13/00, 15/18; C12N 15/00, 15/11, 15/63, 15/67, 15/70, 15/74, 15/79, 15/85; C12P 21/02

## B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

DIALOG (files 155, 5, 73, 357; Medline, Biosis, Embase, Biotech Abs.), APS, GeneSeq, EMBL, Genbank  
search terms: DNA binding protein??, fibroblast, mast cell, hematopoietic, isolat?, purif?, characteriz?, transcription,  
trans-acting, factor, serglycin, proteoglycan, DNA, secretory granule, gene